ARC

January 2019

Occupational Licenses

by James Spector-Bishop and Amanda Rohrer

At the ARC we continue to collect and maintain a central database of licensed occupations, which states are required to submit to us under TEGL. Thank you to everyone who sent us updates this past year. We received and uploaded occupational licensing data from 19 states in 2018 alone. Keep up the good work! Your efforts and cooperation help job seekers navigate licenses in their state through CareerOneStop and help researchers produce meaningful insights about the state of occupational licensing across the nation. As this topic has received more attention more and more groups have expressed an interest in using our information.

While most states submit very comprehensive data, LMI offices are not generally the owners of licensing data and are compiling it from other sources. The data that is available to us is not consistent from state to state (or even department to department within the same state!) and our familiarity with the concepts and subtleties may be limited. The requirements have been pretty loose to accommodate those differences, but that also means it takes a fair amount of judgement for the submitter to decide what goes in the tables they send in. And when those judgement calls are made in each of 50 different states, sometimes decisions are made differently for different areas. This makes using the data as a national data set somewhat difficult.
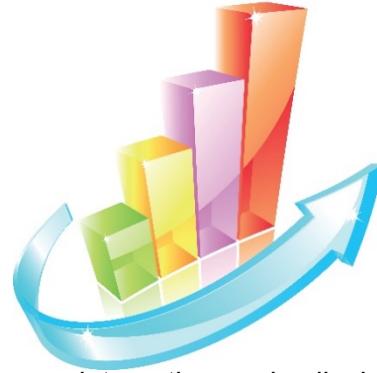
Alternative Data Sources in Tableau

by Amanda Rohrer

Many states use Tableau for data analysis and interactive web display. The software enables analysts to create interactive data experiences without the need for programing experience. Users can connect to a variety of existing data sources; many people start with Excel connection because it's familiar and easy to manipulate, but for a viz that is maintained in the long term Excel has some disadvantages. Among them – typos or copy-paste errors can result in incorrect data in the file, people juggling multiple responsibilities may miss new data releases, and file versions or multiple users can introduce error. Connecting directly to a database source ensures that the viz (and any other data product) remains consistent and up to date.

Tableau allows users to connect to all different kinds of data sources. For many uses connecting directly to the WID (or database views on the WID) solves most of these problems – doing so keeps the source data in a predictable format and prevents incorrect values from showing up in the end product. Maintenance is reduced to keeping track of release dates and refreshing the connection at appropriate times.

There are plenty of times analysts may have reason to use data that's not stored in the database, though. In that case, there's an alternative way to get the benefits of connecting to the database while expanding beyond the data that's internally available. Users can connect directly to publicly available data sources on the internet.

WID API Project

by Gary Sincick

The Workforce Information Database (WID) is a database structure that standardizes how labor market information is stored. It was developed to enable data sharing between states and to facilitate the development of third-party applications that use state labor market information. The WID structure is defined and maintained by the Analyst Resource Center (ARC). As technology changes and state practices evolve, the ARC strives to keep those standards relevant and periodically delves into new areas.

Data Sharing & Database Platforms

The WID structure defines database tables, fields, and relationships independent of any particular database platform (e.g. Oracle, SQL Server, MySQL, etc). However, any application that uses a particular WID database will very likely need to access it in a platform-specific way. This leads to a situation in which states have WID databases that are consistently defined, yet may still have difficulty sharing data between different platforms. Also, vendor applications are usually platform-specific, so the choice of platform may determine the solutions available, and vice versa.

One way to make data more widely available in a platform-neutral manner is to create a layer of web services. These services would provide WID data via http in a standard data format, such as XML or JSON. These web services could provide a foundation for labor market information web applications, or for integrating LMI into other applications.

Most application development now uses web services as the method of choice to access data from a database. This will make application development easier for states, and give them a workable application model that matches the database model.

ARC Activities & Aims

The purpose of this project is to create an API specification for a set of web services for the WID. By working together with interested parties from multiple states, it is hoped that duplication of effort can be avoided, that design issues can be thought through carefully, and a common standard agreed upon.

Once an API specification has been developed, it can be published by the ARC and made available for any state (or other third party) to implement. The goal is for all implementations of the API to be as consistent as possible. There is no plan to require states to create APIs, but this provides a roadmap for states that are moving that direction anyway.

In terms of design, the following items will need to be considered by the WID API workgroup:

- The overall structure and organization of the API.
- The list of tables that will be accessible in the first version of the API.
- Naming conventions for paths, parameters, and data.
- The various functions and methods that the API will allow on these tables.
- How the API will relate to the WID structure.
- How the API and the WID structure will be kept in sync over time.
- Any special requirements for certain tables (e.g. Employer Database)
- What kind of authentication/security will be needed for the API.
- How returned data should be structured (data model)
- Formats of returned data (e.g. JSON, XML, etc)
- How large data sets should be handled (e.g. paging)
- What metadata should be returned
- What http response codes to use
- What (if any) validation should be done on query parameters
- How to handle unimplemented methods (in the case of partial or incomplete implementations)
- If possible, the API should be designed to accommodate third-party tools, such as Tableau

Occupational Licenses *from page 1*

Occupational Coding

In some states there are many specific licenses that apply to a single broad occupation, and in others licenses are listed broadly with descriptions detailing requirements for different tiers or specializations. One example is teachers – some states have a teacher licensure where you're required to demonstrate a specialization in one or more areas, detailed in the application. Others have separate licenses for science teachers, middle school teachers, or music teachers. When users who are unfamiliar with the differences do a rough analysis, this results in an apparent uneven distribution of licensure, even though the practical applications are pretty similar. Rather than restructuring the data (doing a lot of extra work, and losing some of the richness of the existing data), we've started revising the occupational crosswalk (licxocc or licxonet) to standardize how those occupations are treated. So, if there is a general teaching license, that general teaching license will be coded to all teaching occupations, but if there are many more specific teaching licenses, they can be coded to teaching occupations individually. This process is done centrally and reduces the work load on states as the crosswalk to occupations is no longer something they need to spend their time on.

Other Data Sources

A number of other data sources exist where people have attempted to collect licensure information. Mostly they're narrower in scope than what we do, but may contain more detailed information about individual licenses. There's also a data source for professional associations, many of which reference licensing authorities or state licensure to help the people working in their respective fields find the relevant information to legally work. We've sought out these other data sources and can relate them back to the national license database – this allows us both to identify gaps in our data and add richness that may not have been collected.

Documentation

Because historically our emphasis has been on collecting as much data as possible without overwhelming states the guidance for this process has been left pretty open to interpretation. As we analyze data more closely and make note of ways that can vary, we've been expanding the documentation on procedures. While this will likely be helpful to people who are new to the process, it does not mean that states who already have an existing process of equal quality necessarily have to go back and redo everything. We've revised some data centrally and can send that to you, but the goal is to make it easier rather than harder for states. For the most part, that means that we'll identify the changes we make so that we can make them again with new submissions, but if you'd like to keep your database synchronous with the national one let us know and we can send out a revised version of your data.

Table Structure

The table structure that we ask states to submit in can be vague. In many cases, this means we get a lot more information than we would get otherwise, but it makes it harder to use. We're rolling out suggested improvements to the table structure to help filter and use the data. There are two areas we're focusing on – first, the names in the licauth table. Right now there are 3 fields, name1, name2, and name3 – these will be renamed to be more meaningful to the content that goes in them. Second, there will be indicator fields added. Right now all information about the license is put in a description field that has no structure and can't be filtered. A few additional fields will be added to help that process – this may be indicators of the types of requirements (fees, exams, experience, etc.) of the license, or to distinguish between registries, certificates, and licenses. While the description field can remain and contain details, this kind of basic information will help end users compare across states. Most of this information will be obtained from alternate data sources or from the field itself, but in future years we hope to get more accurate information from states where possible.

SOC 2018 Upgrade and the WID

by Jackie Summerton

If you haven't already upgraded to SOC 2018, there's still plenty of time. With WID version 2.7, this is easy to do.

Although 2018 is part of the title, the new SOC codes are not being used yet in the WID core tables. OES started using the SOC 2018 codes for data collection in November of 2018 but OES staff (at least in my state) use BLS's Survey Processing and Management (SPAM) system to collect data. That data will start being loaded to the WID by spring 2020 per https://www.bls.gov/oes/soc_2018.htm/ . Following the publication of the OES numbers in spring 2020, Projections data would then also use the SOC 2018 codes.

On the other hand, there is no real reason to wait to upgrade to SOC 2018 either. Your state may want to implement the SOC 2018 for data outside of the core tables. The full implementation schedule is posted here: https://www.bls.gov/soc/socimp.htm

Everything you likely need is easily downloaded from the Analyst Resource Center (ARC) website from http://www.widcenter.org/document/occcodes/

For the OCCTYPES table, you may need to add a record for the codetype = 19 and codetydesc = SOC 2018 per WID version 2.7 Addenda & errata. http://www.widcenter.org/structure-2/

OCCCODES

🕐 November 20, 2017   👤 Amanda Rohrer

The occcodes table contains industry codes of all types. It references the occtypes table. Depending on what tables your state populates and how long the historical series is, which code types you need may vary quite a bit.

Data Sources

System	Versions	WID Content and Crosswalks	Notes
SOC	2018 SOC	soc2018 soc2018xsoc2010	Not yet in use for publication – the OES program will begin collecting the data in this coding structure with the Spring 2018 panel

This link, http://www.widcenter.org/document_category/core-tables/ has the files you'll need to load to the WID.

The 2018 SOC link under "Versions" will take you to the BLS website with all information about the upgrade. You'll need to download the load files from the links under "WID Content and Crosswalks."

First, the soc2018xsoc2010 link will have you download a file that is ready to load to the OCCXOCC table. This crosswalk from the 2010 SOC to the 2018 SOC matches every detailed occupation from the 2010 SOC with the corresponding new 2018 SOC code(s) and title(s).

When you click on the soc2018 link (under the WID Content and Crosswalks header), you'll find two files to download.

Data.widcenter.org - /download/soc2018/

[To Parent Directory]

01/10/2018	2:22 PM	73469	occcodes.cvs
01/10/2018	3:02 PM	541	readme.txt
01/10/2018	2:22 PM	364742	soccode.cvs

Occcodes.csv is ready to load to the OCCCODES table. Soccode.csv is ready to load to the SOCCODE table.

Follow your plan for loading the files to the WID. The WID documentation, here: http://www.widcenter.org/wp-content/uploads/2017/04/WID27_DataDictionary_20180327.pdf , has a Suggested Load Order in Appendix A.

If you haven't completely implemented the upgrade to WID v2.7 or if you use other tables, (like custom, non-core or deprecated tables), you might need to create other versions of the data and/or crosswalks for your specific needs. Tables like LAYTXOCC (laytitles), LICXOCC (licenses occupations), MATXOCC (microMatrix – Projections), and MOCXOCC (military occupational codes) might need to be updated but currently I have not been able to find a source for these crosswalks. I believe these might be available once SOC 2018 is actually in use in the WID.

In the Spotlight

MATT STEADMAN

Matt works for the Utah Department of Technology Services *as the lead programmer for LEWIS. He is also a member of the Analyst Resource Center Consortium*

When did you join the LEWIS group?

I am new to LEWIS and LMI in general as of April. I have been fortunate enough to work with many wonderful developers and analysts since then to learn all I can about this industry and LEWIS.

That is your current job title?

I have the rather uninformative title of "IT Analyst II." I am the lead developer for LEWIS in the National Systems team in Utah. I work alongside developers for the Projections Suite and the EXPO system.

Are you originally from Utah?

I was born and raised in the Salt Lake Valley. Then I spent about a year and a half in Estonia after high school, before returning to my rocky-mountain home.

What is your educational background?

I have an MBA and a Bachelor's Degree in Computer Engineering, both from the University of Utah.

What is the most rewarding aspect of your job?

My team is the best part of my job. Even though I've only known them for a few months, or less in some cases, I am glad to be able to call them friends.

What is the strangest job you have ever had?

I spent several months working as what was basically as scam-artist. I was honestly terrible at it. No further questions, please.

What is the most frustrating or challenging aspect of your job?

The biggest challenge is the actual work, but that's also what makes it so rewarding. Making sure that LEWIS handles everything just right can be a very tricky thing to pull off. But if it were easy, it wouldn't be so interesting.

What is the most interesting or awe-inspiring place you have been to?

I would have to say it's a tie between Tallinn, Estonia and Zion's National Park. Tallinn's Old Town is considered Europe's best preserved medieval city. It has intricate cobblestone streets and people have lived there for over 3000 years!

What are your interests outside of work?

I'm very interested in the history and art of video games. So much so, I founded a non-profit organization, the Video Game Heritage Institute, to try and document and preserve that history and build a museum/library here in Utah for the cause.

What about your family?

My wife and I celebrated our 12 year anniversary in March, and we have a strong, independent, four-year-old daughter. We hope to have more one day but the first one definitely makes us question ourselves. I'll just say that my two favorite things to do are to spend time with my kid and to spend time away from my kid.

Read any good books lately (personal or work related) that you would recommend to others?

- Maus - Art Spiegelman
- Reality is Broken - Jane McGonigal
- The War of Art - Steven Pressfield
- The Lean Startup - Eric Ries

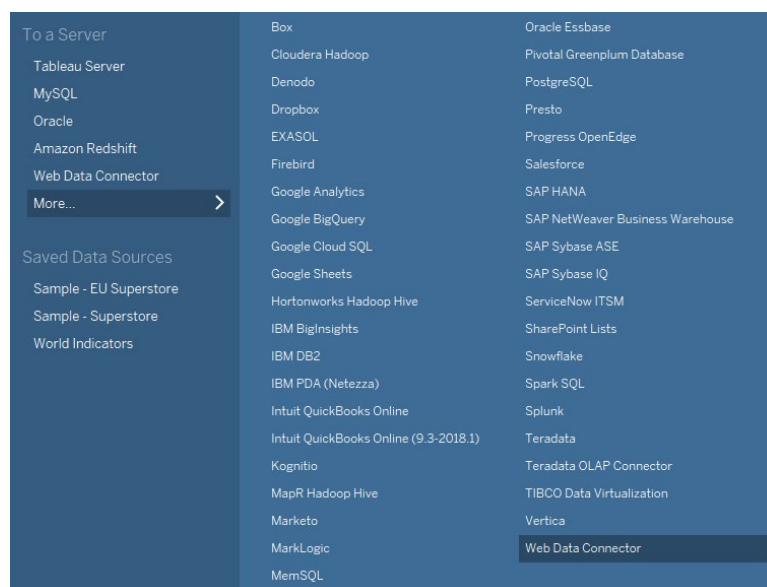### Anything else you would like to share?

I've very much enjoyed working with ARC and my Utah team so far and expect I will be around for a long time yet to come.

Alternative Data *from page 1*

Web Data Connectors (WDC)

Data sources on the internet can come in a variety of formats and structures. They can be XML or JSON, secured or unsecured, single- or multi-table. While Tableau allows you to connect to most types, it needs the structure to be defined before it can do so. The format it uses to do so is called a Web Data Connector (WDC), and is available on the Data Sources screen under more on the "Other Database Connections" section of the menu.

Several of these exist and are [publicly available](#) on the Tableau user forums. That list doesn't include BLS, BEA, Census, or CareerOneStop data connectors, though, which are the large sources of data that are most likely to be of use to Labor Market Information offices.

To a Server	Box	Oracle Essbase
Tableau Server	Cloudera Hadoop	Pivotal Greenplum Database
MySQL	Denodo	PostgreSQL
Oracle	Dropbox	Presto
Amazon Redshift	EXASOL	Progress OpenEdge
Web Data Connector	Firebird	Salesforce
More... >	Google Analytics	SAP HANA
	Google BigQuery	SAP NetWeaver Business Warehouse
Saved Data Sources	Google Cloud SQL	SAP Sybase ASE
Sample - EU Superstore	Google Sheets	SAP Sybase IQ
Sample - Superstore	Hortonworks Hadoop Hive	ServiceNow ITSM
World Indicators	IBM BigInsights	SharePoint Lists
	IBM DB2	Snowflake
	IBM PDA (Netezza)	Spark SQL
	Intuit QuickBooks Online	Splunk
	Intuit QuickBooks Online (9.3-2018.1)	Teradata
	Kognitio	Teradata OLAP Connector
	MapR Hadoop Hive	TIBCO Data Virtualization
	Marketo	Vertica
	MarkLogic	Web Data Connector
	MemSQL	

What does the WDC do?

Some notes about the WDC: All the connector does is define a relationship between Tableau and another data source. It doesn't control either of those – the data still lives and is maintained elsewhere. Server and firewall settings for either the user or the source data can affect the connection and authentication – passwords and user keys – continue to be managed by the data source. Should the structure or content of the data change, the end user still needs to be aware of those changes and change their use of it appropriately.

Once the data structure is defined in the WDC, Tableau reaches out and pulls the data into its own format. It requires an extract for this kind of source, meaning that you're copying it to a local destination and will need to refresh the connection each time there's a new data release.

How can I use a WDC?

The ARC has built a handful of WDCs for connecting to federal data sources containing LMI data. This doesn't cover every useful federal data source and there may be design choices that could be refined. The primary goal of this effort was to create enough different types of connectors to allow states to test the technology and let us know if it's valuable and should be expanded or done differently. This is new territory for the ARC. While the source APIs are pretty stable and changes are not anticipated to be frequent or disruptive, there may be unforeseen maintenance issues. Test them and let us know about issues, and if you consider them valuable let us know that, too – we'll take it under consideration going forward.

Authentication:

Both BEA and BLS APIs require the use of signing keys. These can be requested and are granted in an automated process, but that's something each state will have to get on their own.

BLS: https://data.bls.gov/registrationEngine/

BEA: https://apps.bea.gov/API/signup/index.cfm

There is a sample WDC in the directory that returns earthquake data and can be used for testing before a state has requested a key to one of the others.

Alternative Data *from page 6*

Firewall issues:

Some states have more restrictive policies than others and this process accesses two files on the ARC file server (data.widcenter.org), then passes them through Tableau and hits a federal website. How that's handled state-to-state may vary and if it's insurmountable please let us know!

Tableau version:

During internal testing we discovered that the WDCs require relatively current versions of Tableau. If yours is out of date it may block the connection.

Customization:

Although I've set these up to be available to everyone, you may find your state IT environment or the goals of your office would make it more useful to have your own versions. If you'd like a copy of the files let us know.

Speed:

Sometimes the initial connection is slow. Once the extract is created that is no longer a problem, but it's recommended to start with a small amount of data (a single series, a start date only a few years back) to get a feel for the process.
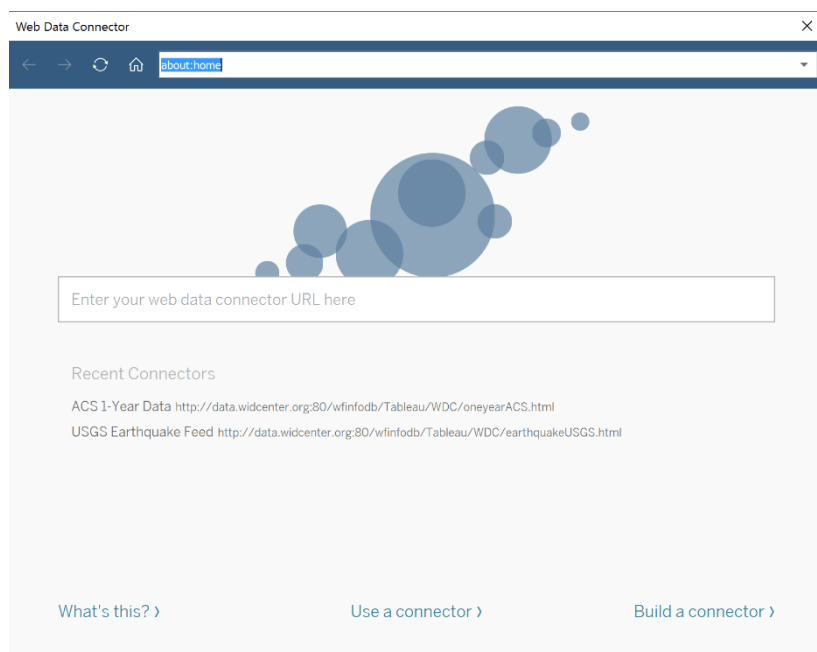
Documentation:

Because the WDC is connecting to a data source that is not controlled by the ARC, information about the data source can be obtained from the source's documentation, which is linked to on the initial parameter input screen of each WDC. There are also notes about what the input fields are expecting on that screen, but if they're not obvious or more clarification is necessary feel free to ask.

Connect to a WDC

Open Tableau. On the initial data connections screen select Connect>To a Server>Web Data Connector. The following opens up:

Where it says "Enter your web data connector URL here" enter the path to the connector you want to use.

Sample USGS Earthquake data (doesn't require an API key): http://data.widcenter.org/wfinfodb/Tableau/WDC/earthquakeUSGS.html

BLS single series: http://data.widcenter.org/wfinfodb/Tableau/WDC/singleseriesBLS.html

BLS multiple series (same as above, but allows comma-delimited list of series ids): http://data.widcenter.org/wfinfodb/Tableau/WDC/multipleseriesBLS.html

Alternative Data *from page 7*

BEA Regional Income: http://data.widcenter.org/wfinfodb/Tableau/WDC/regionalincomeBEA.html

A list of popular BLS series ids can be found here:

https://api.bls.gov/publicAPI/v2/timeseries/popular. You may also already have some saved that you use in the series report query on the BLS website.

Once you've entered the path to the connector, hit enter. You'll come up with a form that's specific to the data source. Enter an appropriate value in each box and hit connect. Once submitted, it will take you to the data connection screen and if you're a frequent Tableau user it should be familiar from there.

Other links:

Look at Tableau documentation here:

https://onlinehelp.tableau.com/current/pro/desktop/en-us/examples web_data_connector.html

For a detailed description of how to connect to a WDC, look here: http://data.widcenter.org/wfinfodb/Tableau/WDC/Connecting to a WDC.docx

BLS Multiple Series   ×

http://data.widcenter.org/wfinfodb/Tableau/WDC/multipleseriesBLS.html

Analyst Resource Center (ARC)

Bureau of Labor Statistics MultipleSeries v2 WDC

BLS Registration Key:

BLS Series IDs, comma separated:

Start Year:

End Year:

Connect

This connects to the US Bureau of Labor Statistics API, please review the BLS documentation

Training

Looking for information that will help you with your job? The Analyst Resource Center has developed a training guide that you may find helpful. Information ranging from *What is the WID?* to *Data Visualization* can be found in this guide. Want information on a specific topic, but don't want to go through the entire training? No problem. You can simply select the section you are interested in learning about. Access to the training guide can be found on the ARC website under the Learning Center tab – Training – click on the ARC Training Guide link.

If there are topics that you would like additional information on, please let us know!

WID API Project *from page 2*

 It is proposed that the WID API be defined using the Open API 3.0 specification (OAS). This is the current industry standard, and is supported by companies such as Google, Microsoft, Oracle, and IBM.

Also, software will be needed to create the actual WID API. This will most likely be SwaggerHub, a cloud-based platform with tools for creating and publishing an API, as well as code generation tools which can be used to assist with API implementations.

Call For Action

Is your state using APIs to retrieve data or power your website? Are you considering a change? If you have relevant knowledge or concerns, please let us know.

The Workforce Information Database is a normalized, relational database structure developed for the storage and maintenance of labor market, economic, demographic and occupational information. The Analyst Resource Center is responsible for the structure development, update, and maintenance of the Workforce Information Database. Current members include representatives: Minnesota (lead), Connecticut, Florida, Iowa, Michigan, Montana, Nevada, North Carolina, Oregon, Utah, Virginia and Wisconsin.

Occupational Licenses *from page 3*

We are here to help! If you are unsure when your state last submitted licenses, or who last submitted them, please don't hesitate to ask us. Also, if you have any additional data which you have collected, we would be glad to have it. We are eager to work with you, so feel free to contact us if you have any questions or feedback!

We can be reached at ARC.DEED@state.mn.us

ARC Newsletter

Editor: Barbara Ledvina

Thank you to Amanda Rohrer, Minnesota; Gary Sincick, Oregon; James Spector-Bishop, Minnesota; Jackie Summerton, Wisconsin, and Matt Steadman, Utah for their contributions to this edition.