Annual Report 2020

Summary of Accomplishments

LEWIS: Software development of a web-based version has gotten approval and made good progress, security hurdles (BLS agent status, LEWIS-ProjectionsSuite interactions, user management) have been addressed

WID 3.0: We've begun the process of setting up a WID 3.0. This will be a major version change, altering some of the field names and lookup tables. The benefits will be a more forward-looking structure that's prepared for future changes and a stronger foundation for our API specification project.

Occupational Licenses: After a long-term project to improve the quality and consistency of occupational license data collected by the ARC, CareerOneStop chose to upgrade their publication tool for the data. We collaborated on the requirements and descriptive content and the new tool is better suited to the strengths of the data. Ongoing data collection and validation processes were fine-tuned to ensure it's possible to keep the data up to date.

Employer Database: The procurement process for the Employer Database has changed and some of the distribution details are being reassessed.

WID APIs: We've been exploring the process for defining a standard structure for APIs based on the WID and comment LMI data. Having these available will give states interested in developing APIs an option for a well-tested structure and thorough documentation. We set up a test environment with functional (but not currently maintained) APIs in 2020.

Documentation: One of the ways we serve states is by documenting potential WID content to help states find a data source that meets their needs. Traffic to our documentation makes up most of our website usage and two topics in particular drive that traffic – CIP-SOC crosswalks and Tableau data connectors. While in both cases we're mostly directing users to other sources, it gives us insight into topics that are of interest to the public and may not be well-documented elsewhere.

Contact list: We've reviewed our contact lists and improved the process for evaluating them. Going forward, we hope this will provide a better foundation for consistent communications and networking.

Contents

The Analyst Resource Center (ARC) is a cooperative group of states that comes out of the Workforce Investment Grant (WIG) funding activities ranging from implementation of the Workforce Information Database (WID), state and regional projections, occupational licenses, to research and publications work. The ARC's primary activities are in setting standards for the WID and in providing resources in support of grant activities that may be useful to a range of recipients. We hold twice-annual meetings with volunteer state participants to make decisions, and most of our resources are published on a website. 2020 was a tumultuous year. Many of our committee members were overwhelmed with new tasks and our traditional avenue of connection, in person meetings, was derailed. Despite this we were able to keep tabs on ongoing priorities and pivot to a digital format for many activities. This report is a summary of those activities and the ones we hope to keep up with going forward.

The ARC exists to help states achieve their data goals, centralizing some of the work of keeping systems up and going. Our efforts often center around the Workforce Information Database (WID), a database structure states implement that allows common definitions and use of data common in the LMI world. Having the common structure:

- Gives states that experience unexpected turnover to have a central resource that can help them make sense of priorities
- Allows improved communications between states
- Allows partner organizations that provide outputs to have a single point of contact for requirements rather than 50 different stakeholders
- Saves state resources from doing design work. Designing a database structure so that it can adapt to changes, include all appropriate details, and documenting that design so that others can use it is challenging. Many states don't have the staff to do that front-end work well and having the structure and documentation ready-made saves on problems that may crop up later

We also procure the Employer Database – a list of employers from a private vendor that allow users to find actual company names (something prohibited by CIPSEA rules for our own products) which would otherwise be available only to a few states, manage the LEWIS system (a piece of software that allows states to securely aggregate OES data for different regions or subsets of jobs than are covered by BLS publication). More recently we've worked on APIs and the challenges states face in either using those provided or in developing their own – again saving the difficult and error prone work of design.

While this type of work is a critical foundation for the management of LMI data in states, design work tends to happen at the front end of a project and the benefits are realized much later. The challenge is in staying ahead of state priorities so we already have structure and taxonomy to offer when states are looking for them and in ensuring that they know those resources are there even if they only need them every several years.

Committee actions

Meetings

Because of travel bans and stay-at-home orders implemented in the early part of the year, our 2020 spring meeting, slated to be in April, was cancelled. A training for states slated to be held in MN in May of 2020 was also cancelled. The uncertainty at the beginning of that time made it difficult to decide if

we should reschedule or plan alternatives, but by mid-summer it became apparent that changing the format was necessary.

For the initial round of meetings to replace the spring meeting, we had four separate 2-hour zoom calls with tight topic-based agendas. While roughly tied to subcommittees in terms of subject matter, all were open to all participants and had good attendance.

At the conclusion of the full-group meetings, it was decided that subcommittees would meet separately before a winter meeting.

Policy had one and Structure, because of the WID 3.0 topic, had 4.

The winter full-group meeting was only one instance and largely addressed the topics of the employer database and WID 3.0.

The notes for the full group meetings are available in [Appendix A](#).

One thing that's missing from the format is the opportunity for states to bring concerns and interest to the group. While the more formal structure we started with for the meetings is suitable when there's clear business to be discussed, there's interest in a more informal meeting structure that would allow more open discussion of state projects. For the spring, we hope to find a time and structure that will work for that.

Training

The cancelled training has not yet been converted to an online format. Between turnover in staff and the heavier workloads many of our trainers are experiencing, organizing that in a cohesive way has been a challenge.

In one of the summer full-group meetings it was decided that the topics should be addressed roughly how they were planned and marketed first to the people who were scheduled to come to the training, but announced in a newsletter for other participants. As follow-up, there may be open invitation forums on topics of immediate interest to states.

WID 3.0

The Analyst Resource Center (ARC) is currently developing a WID 3.0. A departure from the 2.X versions means larger changes to the database structure, especially to the primary key structure. In regular updates, there are fields added to tables, removed from tables, field type expansions (longer text fields, larger numeric types), table deprecations and additions. The core lookup tables are kept the same and field names are unrevised to minimize impact on dependent applications and automation. By contrast, a major version release will change primary key structures, which will affect both lookup tables and data tables. Some core tables may be dropped, and others designated as core. All tables deprecated in the 2.x versions will be deleted.

However, the current structure of the database has begun to create problems as the technology needs of states are changing, and more and more data are becoming publicly available. Looking forward, there's a need to accommodate those changes with a more adaptable database structure. While many tables will remain unchanged, there are some significant improvements we're considering.

- Adding a version number to the areatype concept. Since the WID was first set up the Office of Management and Budget (OMB) definitions of Metropolitan Statistical Areas (MSAs) has changed from an infrequent occurrence to one that's revisited every few years. As a result, accommodating the new MSA definitions is resulting in a lot of new areatypes and they can easily get out of sync as not every state has changes in every new version. By creating areatype categories and vintages we hope to address those problems.

- A minor revision to the time period lookup tables, so that there isn't the redundancy that we now have. Note that this will not affect data table structure.

- Changing field names. When the WID was first created, FoxPro was a major database option, and field names were restricted in length to accommodate its standards. As a result, many of our field names are cryptic or misleading or inconsistent. As we move forward and length isn't as much of a factor in the software states are now using, giving tables and fields more human-readable names will make them more intuitive and make it easier for new users to understand the structure.

- More significant revisions to some of the non-core data and lookup tables will also be considered.

Support

We recognize that this kind of change places more demands on states and intend to provide additional support to aid in the transition. These are some of the proposed offerings:

- Database creation, content, and migration scripts from version 2.8 will be made available as we have for the past few smaller version changes.
- Scripts for views to convert field names and structure backwards to the shorter names – this will allow applications to function as usual without the immediate need for revision.
- A longer-than-typical time to implement changes, written into the TEGL.

Feedback

Our policies for making a major version release are:

1. Structure Committee develops the new database structure.
2. Approved by the ARC Consortium present.
3. Given to ARC member states for comments - 30 days.
4. Revise based on ARC comments.
5. Draft of proposed structure is delivered to states and ETA for review and comment.
6. 90 days for initial comments.
7. Structure Committee reviews all comments after 90-day period; responds as necessary and sends out revised draft of structure.
8. Revised draft sent to all states and ETA for comments - 30-day period.
9. Structure Committee reviews second round of comments and makes final revisions to new structure.
10. Final review by ARC member states.
11. Final structure released, along with supporting documents.

States and other stakeholders will have plenty of opportunities for comment and are encouraged to do so. We would like to hear from the users about any issues that can be addressed, as this is the time to make such changes.

Communications

Given our new, digital format and the switch to similar formats for other committees our members participate in, overhauling our communications and making ourselves more consistent and accessible as a resource is more important than ever before.

Historically, we've maintained a website and a mailing list and while we've attempted forums and Wikis and the use of other tools to connect people none of them have gained much traction.

While we can't necessarily get more engagement from users through tools, we have made some background changes to the website and mailing list that should give us more insights into our users and their interests.

Website

Google Analytics

We use Google Analytics to keep tabs on what pages users are going to on the site itself. The following refer to the time period from Jan 29, 2020 to Jan 29, 2021.

We have no social media presence, so it's a little surprising we got 27 links from that category. Half are coming directly to our site – users have us bookmarked, come from an email, or type in the address. While a significant portion come from searches, many of those are isolated to a couple of topics and often don't stay long.

Referrals are mostly known partner organizations, primarily CareerOneStop by a large margin.
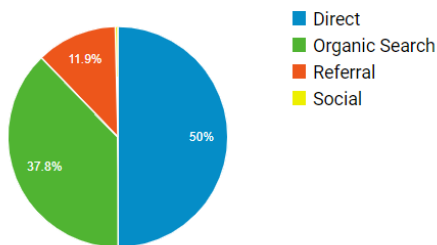
Source	Acquisition Users
	819 % of Total: 11.99% (6,833)
1. careeronestop.org	622 (75.85%)
2. floridajobs.org	48 (5.85%)
3. baidu.com	26 (3.17%)
4. nevadaworkforce.com	23 (2.80%)
5. 10.112.101.182	18 (2.20%)
6. lewissupport.com	18 (2.20%)

Top Channels: Direct 50%, Organic Search 37.8%, Referral 11.9%, Social

Organizational pages describing our tasks and supporting documentation like the TEGL are also a significant driver of traffic.

Selected list of high traffic landing pages	Pageviews	Unique Pageviews
Technical Documentation		
/document/	6230	4078
/document_category/	1312	939
/wid-downloads/	998	582
/workforce-information-database/	493	408
/structure-2/	578	435
/upgrade-the-wid/	236	181
Organizational Documentation		
/arc/	756	593
/im-new/	726	560
/employer-database/	524	398
/state-dbas/	427	342
/tegl/	356	294
/supported-activities/	351	281
/why-a-standard-structure/	326	294
/newsletters-reports/	295	215

One of the biggest services we provide to our users is to provide context and references for where to

Page path level 2	Pageviews	Unique Pageviews
	6,230 % of Total: 17.65% (35,288)	4,078 % of Total: 15.59% (26,160)
1. ☐ /license/	1,182 (18.97%)	843 (20.67%)
2. ☐ /cip-soc-crosswalks/	959 (15.39%)	667 (16.36%)
3. ☐ /all-core-tables/	867 (13.92%)	570 (13.98%)
4. ☐ /occcodes/	512 (8.22%)	306 (7.50%)
5. ☐ /legacy-crosswalks/	473 (7.59%)	260 (6.38%)
6. ☐ /labforce/	256 (4.11%)	175 (4.29%)
7. ☐ /industry/	208 (3.34%)	143 (3.51%)
8. ☐ /ces/	197 (3.16%)	134 (3.29%)
9. ☐ /indcodes/	154 (2.47%)	89 (2.18%)
10. ☐ /iowage/	151 (2.42%)	81 (1.99%)

obtain useful data. That section of the site is a major driver of traffic. The largest landing page is about occupational licenses, which is the only data product for which we are the original source. It's also the one that CareerOneStop uses and may be the destination for their referrals. Our CIP-SOC crosswalk page is also heavily trafficked. The subject is discussed in depth in its own section below. Although we have only ever repackaged the NCES version of this crosswalk for use in the WID, this landing page shows up high in searches of the topic. Core tables describes the deliverables and with the SOC 2018 implementation in OES data this year, occcodes was a major area of interest. Legacy crosswalks are older content referring to no longer active taxonomies. This is used by researchers and occasionally linked to by federal agencies.

File Server

We use logs on our file server to track which files are downloaded. These are set up and managed on the virtual windows server and are extracted each month so that we can parse it into a human-readable format. While there are challenges (when file names change, some links include / at the end, resulting in duplication, there's a record both for a folder and for the asset, meaning that if users navigate

through the file structure rather than being linked directly to the resource the overall numbers can be inflated). This is mostly useful for determining if users are finding experimental resources. It's also important to note that the analytics numbers above do not necessarily equate to downloads of files.

Search Results

We use Google Search Console to track the search terms and landing pages of users finding us through a search engine. While our primary audience already knows who we are and the vast majority of our traffic is direct (from a link or typed into the navigation bar), search results are helpful for determining what topics are of interest to the broader public and what role we could or should have in fulfilling those needs.

Newsletter/Email list

We have an email list with approximately 100 recipients that's mostly used to alert users to new file availability. We also distribute an annual or twice-annual newsletter to that list. In 2020 we converted this from a manually managed list in Outlook to GovDelivery. While users can't add themselves to the mailing list, they can unsubscribe without reaching out. There's also tracking attached to GovDelivery, allowing us to determine which links are popular and how many readers actually open it.

This year only one winter newsletter has gone out, mostly with announcements about the employer database and WID 3.0.

Contact list updates

As part of the employer database updates and the need for better communications, it was decided to combine several separate sources of contacts and make a more universal "technical contacts" list that can be validated regularly. While these may not be people who wish to opt-in to our newsletter, having a contact list for topics such as major WID changes and Employer Database Contract issues is necessary. That's been produced in an initial form. It includes three categories of contacts and allows for any number of interested contacts in each of those categories.

Role	Description
Technical Lead	Person who asks technical questions or details about downloads
Program Manager	Person who asks about schedules, requirements or refers technical tasks to others
Analysis	End user - may have sporadic interactions, but I don't think they have responsibility for the grant

Ideally, it would have at least one Program Manager and at least one Technical Lead because those indicate ownership over certain types of tasks, but while there are contacts for each state, not all of them have a strong interest in the program and some may be outdated.

Projects

LEWIS

The LEWIS application was originally developed in North Carolina as a means to aggregate Occupational Employment Statistics (OES) survey data to non-published regions while still following the very specific

procedures and suppressions defined by the BLS. Because OES data is only published for states and MSAs, many states saw the value in being able to produce reliable estimates for other regions. It also allows wage and occupation filters – some states may use it to produce estimates about minimum wage jobs or combined clusters of occupations. The management of the software was eventually moved to Utah to ensure continuity as its original creator approached retirement and the funding is now managed by the ARC. In 2019/2020 we received approval to begin a new cloud-based version of the application. The original desktop application had many problems – installation in different state IT environments and supporting those differences was cumbersome, security concerns as secure data was stored on a local computer or passwords were shared, and the web-based application has been eagerly anticipated for its easier management.

BLS has approved development in AWS but if it contains any BLS data the GovCloud platform must be used. That environment was quickly set up but there were security hurdles and bureaucratic processes that delayed some steps. LEWIS will likely be up and ready for users to migrate their data to the cloud within a year.

Other updates or topics that the LEWIS team has addressed in the past year include:

- BLS Agent Confidentiality practices: To have access to the data, users need to be BLS agents. The process was discussed with BLS and options for confirming individuals (like use of an API to see current state employees who meet the requirements) were considered. The agreed upon solution is that annually users have to validate their personal email address and an email address that ends in bls.gov as a means of ascertaining confidentiality training.
- Users will have access to their own state's data and cross states' data where there are relevant MSA areas.
- BLS staff have their own level of access for review, including to logs to see who is using the application.
- There are some questions about combining authentication for ProjectionsSuite and LEWIS as both will be hosted in an overlapping environment. Generally, OES data has more security concerns and a team lead role in the application will allow the LEWIS side to review user access and approve requests for data from the ProjectionsSuite software.
- MB3 – the new estimates model – is being reviewed. Matt has been involved in the policy council discussions. The stated policy is still to roll that algorithm change out with the next round of estimates. There is no way the desktop application can replicate the algorithm by next round. MB3 imputes missing data, but since LEWIS creates estimates for non-MSA areas the process would require either using the MSA data they already get to predict rural areas or to get access to QCEW data.

License data

The Analyst Resource Center (ARC) has been collecting occupational licenses from states for several years now. Historically these were compiled in a central location but have only been made available to the public for several years. The publication is handled by CareerOneStop. Download files are made available on the server to facilitate communications with states but they're not advertised on the main site, except in the license guidance.

In 2018 we began a [project](#) to overhaul the license data and try to find what could be validated or collected centrally to make the content more reliable and usable. That was completed with substantial improvements to definitions, consistency of occupational coding, improved structure and the addition of some value-added content that falls outside the normal scope of state responsibilities. Full notes are available in their own [report](#). Much of the last two years has been focused on leveraging that improved content and fine-tuning processes to ensure that it stays up to date.

One improvement was a new CareerOneStop License Finder tool that allows improved searching by occupational code, display of some of the value-added content on the license page, and basic Quality of Life changes. While the CareerOneStop team drove the development process, there was collaboration on the requirements and subtleties of the data between CareerOneStop staff and ARC staff. The result is an application that's more usable and emphasizes the most robust aspects of the improved data set.

In 2020 we also switched from sharing data with CareerOneStop on a sporadic, as-needed basis to having regularly scheduled data releases. These are three times a year, March 1, July 1, and November 1. While the frequency is roughly what it was before, having a firm date makes it easier for CareerOneStop to plan and for states that are submitting data to have a sense of deadlines. There are often edits and revisions or they're curious when they'll see the data on the website, and the regular releases help answer those questions in a more concrete way.

While we collect some data centrally and validate state submissions against other data sources that mean that some parts of the data are always more current than this, the following summarizes the most recent submission we got from each state. With a two-year update schedule, everyone in 2019 or any of the 2020 categories are up to date. I've heard from several in the 2018 and older categories and either already have their new data or expect them with the March 1 release, but there remain a stubborn few where there just doesn't seem to be a person responsible for the task.

Data Release	States Included
11/1/2020	Texas, Washington, Hawaii, Maine, Ohio, Louisiana, Indiana, Maryland, Kentucky, Tennessee
7/1/2020	Colorado, Illinois, California, Delaware, North Dakota, New Hampshire
3/1/2020	New Mexico, Guam, New Jersey, Michigan, Florida
2019	Vermont, Kansas, Utah, Georgia, Rhode Island, Puerto Rico, Alabama, Arkansas, South Dakota, North Carolina, Mississippi, Nebraska, New York
2018	Minnesota, Missouri, Montana, Alaska, Oklahoma, Massachusetts, Oregon, Idaho
Older	Wyoming, Connecticut, Wisconsin, Pennsylvania, District of Columbia, Nevada, Virginia, Arizona, South Carolina, West Virginia, American Samoa, Virgin Islands, Northern Mariana Islands

In 2020 the license data also had some housekeeping tasks, including updating the coding structure from ONET based on 2010 SOCs to their 2019 taxonomy. To accommodate CareerOneStop, which needed to make the switch from 2010 SOC to 2018 SOC all at once in January, November's release included both versions of the codes.

Licensed Occupation Admin Tool

To further improve occupational licenses, the next step will need to be the creation of an administrative tool. Right now a lot of the burden of developing a process to keep licenses updated is placed on individual states. They're expected to do a lot of background prep work and then submit it in a database structure. The skill sets required – designing a process, collecting, editing and reviewing data, then structuring it in a normalized format – are rarely combined into a single position, so many people are pushed outside their comfort zone for this task. While many states have a very good process in place and any tool would allow them to continue that by just submitting the standard file structure, building some of the review process into an application would reduce the burden on LMI shops that are still establishing that process. This will also make the additions we put in centrally more visible to states and give states opportunities to make changes outside the normal two-year review cycle.

With in-person meetings canceled, there may be available funds for development work and we're exploring options. For a high-level overview of the requirements of such an application, see Appendix B.

CIP-SOC Crosswalk

With the release of new CIP and SOC versions, the crosswalks between the two needed to be updated this year. This year like in previous years, the ARC crosswalk was a repackaging of the file made available by the IPEDS/National Center for Education Statistics (NCES), who puts out CIP codes as well.

Because of the nature of CIP codes and the fluid interactions between educational background and destination occupation, a crosswalk between the two taxonomies is never going to be universal. Sometimes the goal is to know the actual labor force outcomes of graduates, regardless of whether the specific coursework of their degree applies. Other times you're looking for the target of the educational program – what occupation is it aiming to prepare graduates for? Sometimes you want every possible match and others you want a very narrow focus or a best match. For many programs, the outcomes can be dramatically different based on the degree obtained – programs can offer a masters, bachelors, and associate's degree option and the career students are prepared for branches significantly.

Rather than attempt to either make a crosswalk that can be applied to any problem or to make many versions for different possible applications, we've continued producing a file in the WID structure according to past practice (using NCES data), while actively seeking out other publicly available options and documenting them as an option for our users.

Looking at our search traffic from the Google Search Console dashboard, the CIP-SOC crosswalk information is one of the primary means that internet users find us in a search. Our content is highly specific and the vast majority of overall website traffic comes from direct navigation (people who type in the URL or click a shortcut), but several CIP variations show up in the list of search terms that result in users clicking on our site. While the raw number of clicks is low, it's important to note that the position for all the variations of CIP-SOC crosswalk is low – the topic itself is niche, but we appear in the first page of results.

CIP-Related search terms that led users to WIDCenter.org, 2020

Query	Users who landed on our site Clicks	Appearances in the search results Impressions	Percentage of impressions that resulted in a click CTR	Rank in the returned results – lower than 10 is the first page Position
cip to soc crosswalk	73	1298	5.62%	4.83
cip soc crosswalk	39	761	5.12%	4.59
soc to cip crosswalk	17	280	6.07%	5.00
soc cip crosswalk	12	134	8.96%	4.25
cip crosswalk	7	235	2.98%	5.83
2018 cip to soc crosswalk	7	30	23.33%	3.33
cip to soc	6	188	3.19%	4.99
cip 2020 crosswalk	4	164	2.44%	7.19
cip code crosswalk	2	69	2.90%	9.42
cip and soc codes	2	22	9.09%	4.18
cip to soc crosswalk table	2	15	13.33%	4.13
cip code crosswalk 2020	1	44	2.27%	7.41
cip code lookup	0	69	0.00%	37.81
cip 2010	0	54	0.00%	52.69
cip codes 2010	0	51	0.00%	73.82
soc cip	0	20	0.00%	4.95
cip table	0	19	0.00%	10.53
ipeds cip codes	0	19	0.00%	77.68
cip soc	0	14	0.00%	4.71
cip codes crosswalk	0	11	0.00%	6.00
2020 cip code crosswalk	0	10	0.00%	7.10
code cip	0	10	0.00%	84.30

The traffic from CIP-related searches seems to be pretty consistent through the year, with no seasonal spike corresponding with fiscal, calendar, or school year. The chart for the most common search term (cip to soc crosswalk) is below. The blue line are people who landed on our site, purple are the ones who saw our website in their results.

	Total clicks	Total impressions	Average CTR	Average position
	73	1.3K	5.6%	4.8
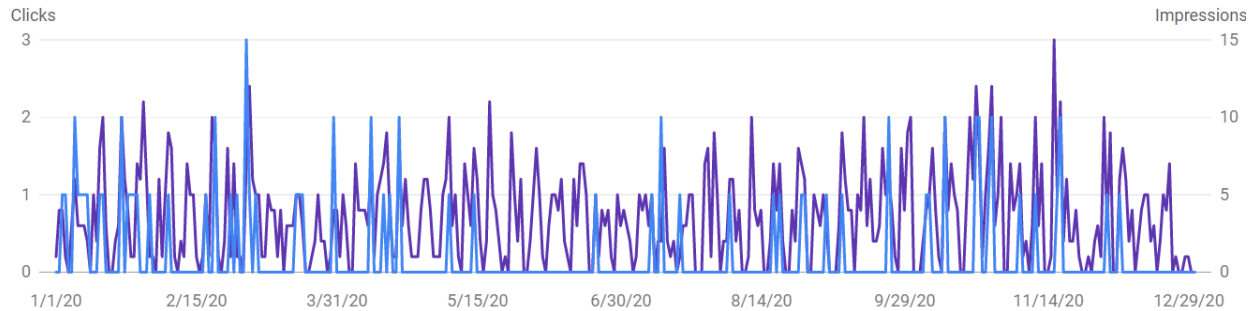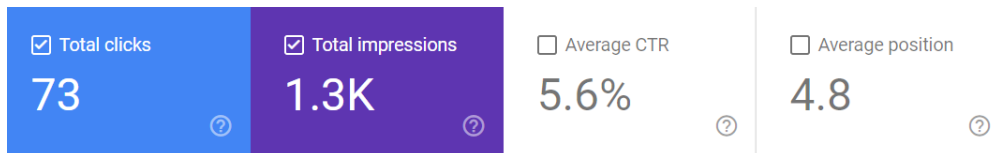
Other than the download files themselves, we really only have the one page devoted to the CIP-SOC crosswalks. The pages the searchers land on reflect this as the major driver of search traffic. The cip-soc-crosswalks page shows up both 2nd and 6th in search landing pages for 2020. The duplicate is a result of implementing a security certificate in March that changed the URL to https.

Pages on WIDCenter.org found through search, 2020

Page	Clicks Users who landed on our site	Impressions Appearances in the search results	CTR Percentage of impressions that resulted in a click	Position Rank in the returned results – lower than 10 is the first page
https://www.widcenter.org/tableau-wdcs/	382	32299	1.18%	34.07
https://www.widcenter.org/document/cip-soc-crosswalks/	301	6729	4.47%	11.11
https://www.widcenter.org/workforce-information-database/	259	6017	4.30%	24.71
https://www.widcenter.org/	199	2767	7.19%	25.57
https://www.widcenter.org/employer-database/	186	2918	6.37%	29.17
http://www.widcenter.org/document/cip-soc-crosswalks/	176	2878	6.12%	14.68
https://www.widcenter.org/tegl/	156	4164	3.75%	6.31
http://www.widcenter.org/employer-database/	115	1419	8.10%	33.71
http://www.widcenter.org/workforce-information-database/	112	2734	4.10%	26.63
http://www.widcenter.org/tableau-wdcs/	100	9860	1.01%	31.51
https://www.widcenter.org/why-a-standard-structure/	99	2838	3.49%	12.65
http://www.widcenter.org/	95	2127	4.47%	28.97
https://www.widcenter.org/document_category/crosswalk-tables/	83	4730	1.75%	16.18
https://www.widcenter.org/2019/05/02/apis-and-tableau/	76	1670	4.55%	33.89

Traffic to our documentation (not search results – all users) is concentrated in the documentation feature as noted previously. Among that, the cip-soc-crosswalks page is the second most popular for pageviews with nearly a thousand over the last year, again with no obvious seasonal pattern.

Because of our position in the search results and our historical role in distributing this crosswalk (prior versions of the IPEDs crosswalk were more difficult to find and may have been less accessible or timely

until a few years ago), the ARC has a certain amount of name recognition on this topic. Maintaining this landing page, seeking out new options for reliable crosswalks, and getting both our state users and other interested parties to the best expert option should remain a priority.

Employer Database

Job seekers and economic developers often have a need for a list of businesses by region or industry as a contact list. LMI data is protected by CIPSEA and no non-aggregated individually identifiable data can be shared even when the details are no more than what could be found in a phone book. To fill this need there are a number of private companies that offer a range of products, often marketed to libraries and firms building client sales lists. The products are expensive and their off-the-shelf options may not suit state purposes so a joint effort to procure this product for all states was initiated to negotiate better rates, specific deliverables, and to enable even states with more limited economic development budgets to have access.

The contract has been run through ARC for 24 years and in that time there have always been challenges around state procurement. Individual states have been contract holders, and all have been sued by the company that did not receive the contract, creating a costly legal process. In early 2017 ARC approached ETA to procure this through the GSA schedule. The RFP process was allowed to lapse and there were challenges along the way, but the contract with InfoGroup (now DataAxle) was executed on April 3, 2020.

Looking forward, we're considering changes to distribution. Most immediately, a discontinuation of mailing DVDs to each state. An expansive contact list has been developed and an email sent to ARC contacts, DataAxle distribution contacts, and LMI Directors notifying states that the DVDs will not be delivered beginning April 1. The contract does specify that as a deliverable so depending on the response there may need to be short-term exceptions.

DataAxle also has a newer and more adaptable means of distributing data files than the MFT system they use for states. Conversations have begun and we've agreed to test that system against our current product in the April deliverable window. Accepting this change would require a change to the database structure – while the timeline is such that we could use our normal processes to implement those changes, it will require documentation and communication with states to ensure consistent application. The DataAxle staff also noted that they filter the data according to our requirements – it's possible that the presence of closed businesses may result in a level shift and require clear explanation on how to use the revised data.

Exploratory Projects

The ARC is in a unique position of needing to know what the next priority of states will be to design database structures and documentation to support those efforts. While our members' connections to other advisory bodies and programs can help track those future efforts, there's also a need to follow topics that are outside government and occasionally experiment with new offerings.

WID APIs

The idea of the WID API is to create a structure that would provide flexible access to core WID tables. In the same way that the WID structure is a standard, the API would be a standard that could be adopted and implemented by states as they wish. The philosophy was to be closely tied to structure so that

concepts and conventions would be familiar. It's also database oriented rather than application oriented, which is the more typical approach to API design but is not very flexible for our uses. Building an API or a database is technically not very difficult. Designing one so that it can adapt to changes, include all appropriate details, and documenting that design so that others can use it is much more challenging. By offering up both structures to states we do work that would either take significant state staff time or would be done incompletely or with errors from inexperience or haste.

In considering API options we looked at existing standards for return type and metadata to design a broadly applicable and coherent standard. It's now on SwaggerHub and is publicly viewable. It's also on Basecamp for group communications which hasn't been updated in a while because work has slowed down, but the documents and files are useful.

One problem was that we wanted a live version of the API for states to test but the Oregon arrangement did not allow external access. Instead, a sandbox environment was created in our existing AWS services – api.widcenter.org. It's still necessary to correct some issues created by moving from Oracle to MySQL as the back end, but it will be ready soon for states to test.

The API itself is database agnostic, and the consumer of the API doesn't care what it's implemented against, but to create it you have to point to a database. We are using MySQL for cost and because they're moving toward it in Oregon. One specific issue is how paged results are handled without a standard SQL function. Identifying those issues and how they would be addressed in different implementations is a next step. Some of the features we add are more technically challenging but add value – as we test it, we can determine if a simpler implementation is valuable for states working off different platforms or if the metadata we attach is important. Those are discussions that a working group can have once people can work with the API.

Tableau WDCs

To connect to APIs via Tableau, the connection properties need to be defined in what's called a web data connector (WDC). It's basically an HTML form that collects parameters (authentication information, years or variables) and passes them to a JavaScript application that creates a URL and GET call and defines the structure of the output so that Tableau knows the format it will be displayed in. Because they're tiny applications, they have to be hosted on a server – it becomes more complicated to keep them on the same local computer that you're using to connect to Tableau. In the process of trying to help users get BLS and Census data via API, I discovered that Tableau WDCs, which are relatively new, are not generally provided by the API creator. For my own use and other states that don't have developers available to create and maintain them, I put out a few limited use ones for BLS and Census data. Despite a few hiccups (Census API formats changed in a small but significant way resulting in null data sets for a while and some non-state users have requested they be restructured to return multiple years or smaller area types), the existing experimental WDCs have been a driver of search traffic, though very little of it is specific to the sources I was interested in (Census and BLS).

WDC or Tableau related search terms that led users to WIDCenter.org, 2020

Query	Clicks Users who landed on our site	Impressions Appearances in the search results	CTR Percentage of impressions that resulted in a click	Position rank in the returned results – lower than 10 is the first page
tableau web data connector list	17	585	2.91%	9.13
tableau web data connector	3	1655	0.18%	28.61
wdc tableau	3	155	1.94%	10.27
tableau web data connectors	3	62	4.84%	19.47
tableau wdc	2	296	0.68%	11.44
tableau census data	2	41	4.88%	17.05
wdc connect	1	1090	0.09%	5.37
web data connector tableau	1	173	0.58%	30.18
web data connector tableau server	1	29	3.45%	27.79
tableau json web data connector	1	20	5.00%	20.15
create web data connector for tableau	1	2	50.00%	29.00

The number of clicks that get to the landing page for that topic isn't much higher than the CIP-SOC-crosswalks, but there's a much greater variety of searches that are seeing us in the result and many more impressions. We're rarely on the first page of results, though.

Pages on WIDCenter.org found through search, 2020

Page	Clicks Users who landed on our site	Impressions Appearances in the search results	CTR Percentage of impressions that resulted in a click	Position Rank in the returned results – lower than 10 is the first page
https://www.widcenter.org/tableau-wdcs/	382	32299	1.18%	34.07
https://www.widcenter.org/document/cip-soc-crosswalks/	301	6729	4.47%	11.11
https://www.widcenter.org/workforce-information-database/	259	6017	4.30%	24.71
https://www.widcenter.org/	199	2767	7.19%	25.57
https://www.widcenter.org/employer-database/	186	2918	6.37%	29.17
http://www.widcenter.org/document/cip-soc-crosswalks/	176	2878	6.12%	14.68
https://www.widcenter.org/tegl/	156	4164	3.75%	6.31
http://www.widcenter.org/employer-database/	115	1419	8.10%	33.71
http://www.widcenter.org/workforce-information-database/	112	2734	4.10%	26.63
http://www.widcenter.org/tableau-wdcs/	100	9860	1.01%	31.51

The actual users of the tools are displayed monthly below. Note that the WDCs are pairs of files – the HTML collects the parameters and then if they click submit on that form, the .JS actually runs the query. Use is limited but consistent. The earthquakeUSGS API is well-established and doesn't require authentication, so it's an example used to see how they work before you're issued an API key. It's not interesting to our target audience and there are many more useful applications pointed at it – the consistent numbers of users might indicate a baseline of how many of the users that land on our page are trying to understand what a WDC is as opposed to looking for one to use regularly.

WDC Usage

Cs-Uri-Stem		Moyear										
	2-2020	3-2020	4-2020	5-2020	6-2020	7-2020	8-2020	9-2020	10-2020	11-2020	12-2020	1-2021
/wfinfodb/Tableau/WDC/singleseriesBLS.js	27	5	26	67	82	11	34	19	13	33	14	5
/wfinfodb/Tableau/WDC/singleseriesBLS.html	33	8	29	41	75	10	35	19	16	33	14	5
/wfinfodb/Tableau/WDC/acs2020revision.html		145	45	15	18	26	17	10	12	10	42	10
/wfinfodb/Tableau/WDC/acs2020revision.js		144	42	17	19	25	16	8	11	10	42	10
/wfinfodb/Tableau/WDC/multipleseriesBLS.js	10	3	42	88	62	35	10	15	10	21	13	14
/wfinfodb/Tableau/WDC/multipleseriesBLS.html	13	10	39	59	56	36	13	17	12	22	10	14
/wfinfodb/Tableau/WDC/earthquakeUSGS.html	17	16	23	12	13	11	27	25	8	13	20	4
/wfinfodb/Tableau/WDC/earthquakeUSGS.js	11	14	20	16	15	10	12	21	7	13	20	4
/wfinfodb/Tableau/WDC/variablesACS.html	21	17	15	9	5	7	6	4	11	10	8	4
/wfinfodb/Tableau/WDC/regionalincomeBEA.html	14	11	14	15	20	5	6	5	6	6	5	3
/wfinfodb/Tableau/WDC/regionalincomeBEA.js	7	7	11	15	30	4	4	5	5	6	5	3
/wfinfodb/Tableau/WDC/variablesACS.js	11	14	13	9	4	6	5	4	10	9	8	4
/wfinfodb/Tableau/	26	13	5	9	8	3	5	5	3		4	1
/wfinfodb/Tableau/WDC/	10	13	4	5	5	1	5	2	1		2	1
/wfinfodb/Tableau/WDC/oneyearACS.html	12	1		10			1	1				2
/wfinfodb/Tableau/WDC/oneyearACS.js	1			19			1	1				

WDCs are heavily dependent on the APIs they're directing users to, both in terms of structure and content. In the Census one, for example, areas smaller than a county require an additional parameter in the search string, essentially forcing you to create two possible connections and have the WDC choose which you want if you're going to allow small area selections. Add to that the complication of collecting human-readable inputs and converting them to Census codes, the WDC gets larger, more complicated, and potentially slower the more it can return. Same with years – the Census API doesn't allow multiple years, so to return results for different data sets means that in the background the WDC would have to recognize the multiple data sets and then choose to run multiple times and concatenate.

The Census Bureau produces a massive amount of data and their API is correspondingly complex. Given the annual update schedule of most of their products downloading files off data.census.gov is not that cumbersome. For internal uses, having the WDC connection is just not that helpful. However, the search results and the number of users who are digging through several pages of results to get to widcenter.org suggests that there is a demand for this type of tool and meaningful information about the topic. WDCs would be better designed and maintained by the creators of the APIs.

Non-BLS areas

BLS data products generally produce data for states, MSAs, and sometimes counties. However, states often will aggregate county or city data to form different regions which may be more meaningful for economic development. This is particularly important regarding MSAs – in the past decade, the MSA definitions for Minnesota have changed three times and have expanded beyond what most industry professionals would consider the core of the Twin Cities. The newest counties added to those definitions are different in economic and political characteristics – rather than being inner ring suburbs with residents who work, entertain, and send their kids to school in the urban cores of Minneapolis and Saint Paul, they're exclusively commuters who don't consider themselves part of the Cities in the same way. Similarly, the balance of state areas that are defined mostly by not being in a metro are not uniform in their economic characteristics. Southern MN is largely agricultural with a few anchor manufacturers, while Northern Minnesota relies on mining, logging, and tourism. It's not meaningful to group them together for most analytical purposes. Many states are in the same boat – the criteria they use for defining geographic areas needs to be different and more stable than current MSA procedures allow. As such, they often define their own.

Workforce Development Boards (WDBs)

Workforce Development Boards are state-defined regions for distributing federal Workforce Innovation and Opportunity Act (WIOA) funding. CareerOneStop already collects and maintains information about WDBs. For the past couple of years, we've taken their information and restructured it for the appropriate WID tables, as well as produced appropriate shape files for geographic use. Those aren't distributed except by request because WDBs often combine area types, which means you have a mix of counties, cities and balance of counties. The WID-format GEOG and SUBGEOG tables have been downloaded from the server 50 times since they've been maintained.

Economic Regions

There's another category of region that's largely county-based and tied to state requirements rather than federal funding streams. In Minnesota there are two levels of aggregation and they're called Planning Regions (larger) and Economic Development Regions (smaller). Many other states have the same concepts but not always labeled the same way, and more relevantly, their LMI offices aggregate existing BLS data sources to produce and publish data for those regions. In the early part of 2020 we went through the state LMI pages looking for the types of regions for which OES data is published. The details of that review are contained in Appendix C.

Of the 52 states (including Puerto Rico and Washington DC), 30 produce OES data for non-BLS areas, and most of them make it available for download on their public websites though some are only totals or contained in reports. Clearly, non-MSA regional breakouts of states are important to many LMI shops. However, updating the initial survey of areas is probably not reasonable even annually. The new LEWIS application requires that users flag "publication" data so that any future aggregations won't violate confidentiality inadvertently. As part of that, the definitions of areas are stored. Without sharing any of the actual data, it should be possible to get an extract of regions and their component areas from the LEWIS team annually. This is likely to require substantial cleanup even so but knowing what regions state experts consider valuable has the potential to improve the way national data providers like CareerOneStop display and filter data and could help inform future MSA definition methods. With the most recent Federal Register Notice significant definition changes to MSAs are coming in 2023 and they're anticipating further discussion for the 2030 Census and having this as a resource could be helpful.

Real-Time analytics

Real-Time analytics of hiring has been a topic of interest to states for several years now. Generally, the term applies to tools provided by private companies that attempt to describe hiring practices based on the information available on the internet. These are either from companies that own websites that are job seeker destinations and host a large number of advertisements on their own, or companies that scrape a pool of major employers to develop metrics. Because of the nature of the source data, the sampling methods aren't traditional and among industry experts the statistical value of real-time analytics based off job postings is debated. Despite this, states remain interested in the topic because they fill a niche that our existing data sources can't. While a few states have a Job Vacancy Survey to look at hiring, most do not and even those that do have infrequent surveys. Real-time analytics are more current than any of our employment statistics – they can be up-to-date to the day, compared to monthly for LAUS and CES. They also are not subject to suppression in the same way and their

cost/sample is much lower. At least in theory, this means it's possible to get much more granular data, particularly in rural areas where Labor Market data is scantly available.

Unfortunately, evaluating the quality of the products available has been a challenge. They're expensive, sometimes report as percentages or indexes that don't compare easily to other sources, and their methods are deliberately obscured to protect the companies' intellectual property.

During 2020, several of the companies that provide real-time analytics released free versions of their products to aid in the pandemic efforts. The ARC also initiated conversations with the National Labor Exchange (NLX) to get a daily file of the job advertisements they have active and those are currently being loaded into a database daily so that it can be used as a time series of postings with all described characteristics and terms open filterable. This has made it possible both to compare the trends of multiple sources and to look at the more specific challenges faced by trying to develop a trend out of a set of advertisements.

Comparing Job Posting Trends from Available Sources
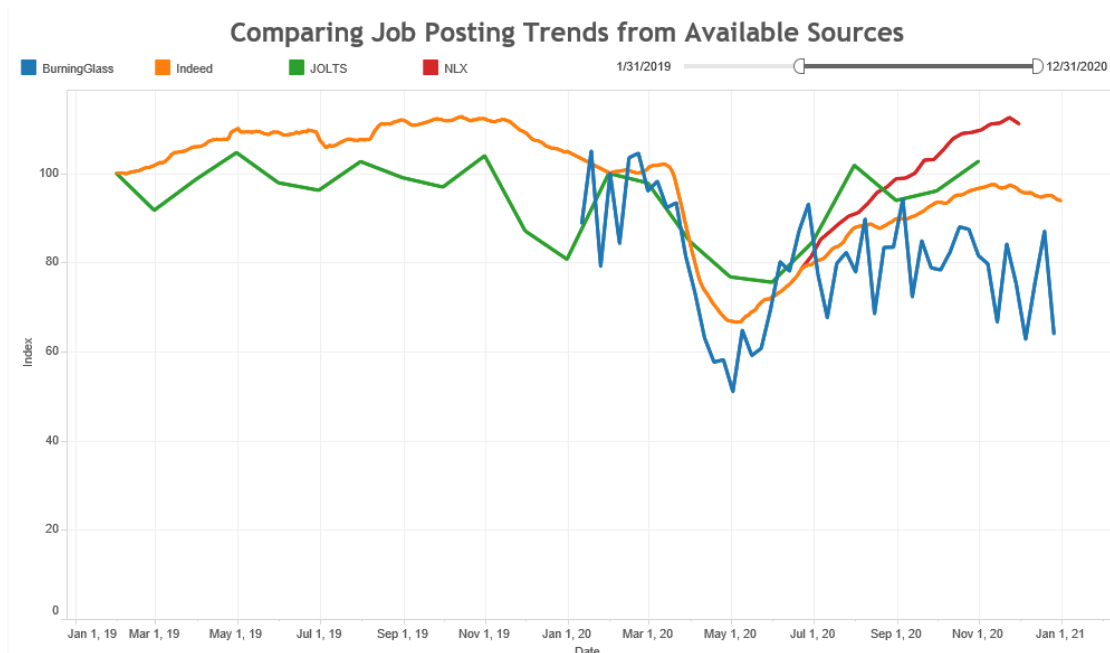
Conversations with NASWA and NLX have started and in the coming year hope to keep a dialog going about the methods and cleanup of that as a data source. If it ultimately can be made available to states like the employer database and the methods are more transparent and well-documented than existing products, it could be a benefit to state LMI offices. With our current microdata access, we're in a position to document and understand the limitations of the source. Topics that our staff may look at are outlined in Appendix D.

Future Directions

There are a number of topics that come up in questions or when designing applications that intersect with our strengths and could be pursued.

Credentialing: While ARC's responsibility is only occupational licenses, CareerOneStop collects and uses professional organizations and certifications. Apprenticeship information is also out there in various

forms. There are other organizations doing a lot of work on the subject now. Keeping tabs on the progress made in this area and helping states find the resources available to them is an ongoing priority.

Non-US data: Recently I've gotten two separate requests about a crosswalk from SOC to ESCO/ISCO occupational codes. A crosswalk from SOC 2010 to ISCO is available on the BLS site, but hasn't been updated with the new taxonomy. Finding out who does this or how it's done could be a priority.

National UI data: While ETA makes a lot of data available on their site, some states have looked at making comparisons to their neighbors based on that published series. Unfortunately, the documentation seems to reference UI concepts that aren't intuitive and it's not always easy to determine what the best metric to use in a comparison would be. Offering some guidance and compiling a list of resources would be timely.

Employment Training Provider lists (ETPL): Current WIOA requirements involve the publication of an approved provider list for employment training for each state. That's somewhat fragmented, but is becoming more centrally available. We already have a significant number of our users with an interest in IPEDS data formatted for the WID, keeping tabs on ETPL progress in case there's a resource that needs to be documented or promoted to our users should be a priority.

Relationship-building: NLX, CredentialsEngine, federal efforts. Participating in comment periods for data sources.

Entry Level Jobs: Many of our applications start at a selection list of the whole SOC taxonomy. Searches are always improving and might make them a little better, but it's still an overwhelming amount of choices for job seekers. Using ONET characteristics and other research, I've done some preliminary work on distinguishing between low-skilled jobs with little room for advancement, entry-level jobs with potential pathways for advancement, specialized jobs, and entry-level but with a high education requirement. With a systematic and defensible methodology, having a category that people understand rather than obscure formulas of wage and demand power our selection lists could make labor market tools more usable.

Appendix A Meeting notes

LEWIS (led by Matt Steadman): 7/23/20 at 11 Central via Zoom

AWS Updates:

The annual major of LEWIS was successfully put out. Temporary SOC hybrid code definitions are in place, but there were some last minute code changes from BLS that created some hiccups. The output file is still using the codetype 14 (SOC 2010).

BLS has approved development in AWS but if it contains any BLS data the GovCloud platform must be used. That environment has been set up but security hurdles have so far blocked access by developers. Progress can be made in the interim provided no actual BLS data is used. Matt noted that some of the problem is created by working under the umbrella of Utah IT which also adds cost to the platform. If the problems persist setting up the GovCloud platform under ARC may be possible but it cuts off access to Utah IT experts.

Introductions:

Venugopal (Venu) Japa has been a part of the development team for a few weeks now and has so far familiarized himself with the software as a user and is now moving on to understanding the inner workings of the application.

Patrick Flaherty (Assistant Director of Research and Labor Market Information in CT) attended because of role in LEWIS and expected turnover.

Steve Williams (Oregon) also sat in.

Future developments:

LEWIS will likely be up and ready for users to migrate their data to the cloud within a year.

MB3 – the new estimates model – is being reviewed. Matt has been involved in the policy council discussions. The stated policy is still to roll that algorithm change out with the next round of estimates.

There is no way the desktop application can replicate the algorithm by next round. MB3 imputes missing data, but since LEWIS creates estimates for non-MSA areas the process would require either using the MSA data they already get to predict rural areas or to get access to QCEW data. The AWS platform would help solve these problems, but the existing desktop application can't easily.

The question was raised if BLS has weighed in on whether it's allowable for LEWIS to keep producing estimates using the old algorithm after the OES transition. For internal purposes it's fine, and offering the option after the change is implemented is also probably fine, but no final verdict about not implementing the change immediately. Matt will follow up on Monday and LEWIS already does many things that OES does not, so there is precedent.

Next meeting

There were some connection problems using Zoom, but it was decided not to switch platforms yet. Utah can host in GoogleMeet, Microsoft Teams works for some states.

The group decided to start a workgroup to monitor LEWIS progress, meeting monthly and likely continuing through the roll-out of the cloud version of LEWIS.

_____

Employer Database (led by Steve): 7/28 at 1 Central

In attendance :Amanda Rohrer, Steve Duthie, John Pearce, Bill McMahon, Joe Jaehnke, Mike Peery, Steve Hine, Patrick Flaherty, Patrick Flaherty, Matt Steadman, Christopher Robison, Dana Placzek, Gary Sincick, Kevin Doyle, Steve Williams, Robert Bean, Rajani Edupalli, Bob Viegas, Tim Kestner, VenuGopal Japa, Al Sylvestre

Background

History of Employer Database – the contract has been run through ARC for 24 years and in that time there have always been challenges around state procurement. Individual states have been contract holders, and all have been sued by the company that did not receive the contract, creating a costly legal process. In early 2017 ARC approached ETA to procure this through the GSA schedule. So the RFP process was allowed to lapse, There were challenges along the way, but the contract with InfoGroup was executed on the third of April.

Terms – 1 year with 4 optional extension years (locked in through 2024 deliverables)

Cost – $410,000 for each year (fixed price), that's a 10k increase for each year, but we'll also need to consider how to handle support to states and what that needs to include.

Details and Requirements of the Contract

The language of the requirements is much more general than previously detailed in the CT contract. That raises some questions about how we go about managing this arrangement and what the ARC's role is in that.

Are there details or requirements that are contained elsewhere that we should know more about? One example – In clause 2c….Mandatory Product Requirements, the current language only specifies that states can query by industry, while the previous had requirements about searching on 2-digit NAICS and whether it needed to be limited geographically. There were limits to the number of records displayed at a time and types of mapping that we also need to include for state users. Despite the more detailed prior contract, there were also a lot of things that needed to be defined for specific use cases before, and many of them are in COS technical specifications. While we want to protect InfoGroup's data, we're unclear on what restrictions we can define and whether enforcement then falls to us.

Some discussion about eligible sublicensees occurred and states noted that it was cumbersome before and often avoided, but each had their own approach.

We need to organize our requirements in a way that we detail this and provide it as a document to state subgrantees, if only for their convenience. Would that be legally binding? Lack of strict language in current contract might make it easier for spurious lawsuits to gain a foothold.

Someone from the ARC will work with Bob (Bill and Steve, Al) to deal with the detail. Specific topics may include:

- Search parameters
- InfoGroup logo
- Mapping, number of records displayed
- Sublicensees
- Captcha
- Enforcement

Have there been any issues with state access to the database since April?

- NV had a few problems,
- there were some password change issues for a week or so.
- Overall a pretty seamless transition

How can the ARC fulfil the requirements of the contract?

There are different approaches that could be taken – a standing subcommittee, an individual who can be specifically responsible. The ARC might serve as the point of contact and forward requests on from there.

One of the old problems we had was monitoring states' compliance with the contract. Several would be out of compliance and then we'd have issues once it was discovered. This role should be defined as we assign our contract support.

CareerOneStop Business Finder already has these rules applied correctly and promoting their widget or API as a way to meet the TEGL requirements could both aid states in their obligations and reduce enforcement issues.

Might be useful to survey states to see what kind of demand there is for intermediate user access.

We'll have another meeting, conversations with Infogroup.

_____

Future Projects (led by Bill and Gary): 7/29 at 12 Central

In attendance: Joe Jaehnke, Gary Sincick, Amanda Rohrer, Nicole Kennedy, Steve Hine, John Pearce, Steve Duthie, James Spector-Bishop, Dana Placzek, Kevin Doyle, Mike Peery, Matt Steadman, Bill McMahon, Steve Williams, Christopher Robinson, Tim Kestner, Patrick Flaherty, Mike La Rue, Don Haughton, Al Sylvestre,

API Planning

Quick review – The idea of the WID API is to create a structure that would provide flexible access to core WID tables. In the same way that the WID structure is a standard, the API would be a standard that could be adopted and implemented by states as they wish. The philosophy was to be closely tied to structure so that concepts and conventions would be familiar. It's also database oriented rather than application oriented, which is the more typical but is not very flexible for our uses.

Looked at existing standards for return type and metadata to design a broadly applicable and coherent standard. It's now on SwaggerHub and is publicly viewable. Also on Basecamp for group

communications which hasn't been updated in awhile because work has slowed down, but the documents and files are useful.

The advantage to building an API specification is that we can both keep it close to the WID but it doesn't need to be exact – can improve field names or other inconsistencies without creating extra work for states.

One problem was that we wanted a live version of the API for states to test but the Oregon arrangement did not allow external access. Matt created a sandbox in our existing AWS – api.widcenter.org. It's still necessary to correct some issues created by moving from Oracle to MySQL as the back end, but it will be ready soon for states to test.

The API itself is database agnostic, and the consumer of the API doesn't care what it's implemented against, but to create it you have to point to a database. Java vs .NET, etc. We are using MySQL for cost and because they're moving toward it in Oregon. The specific issue is how paged results are handled without a standard SQL function. Identifying those issues and how they would be addressed in different implementation is a next step. Some of the features we add are more technically challenging but add value – as we test it, we can determine if a simpler implementation is valuable for states working off different platforms or if the metadata we attach is important. Those are discussions that a working group can have once people can work with the API.

We will set up a working group with regular meetings.

NLX potential and grants

CareerOneStop has many APIs. The initial approach was just to expose the back end of the applications they build and let other parties use them, but they've expanded in their use and many more customers want to do more or different things than COS applications. Now as new ones are built COS is looking beyond their application-specific APIs and trying to plan for other uses.

The most popular are the things that you can't get elsewhere, like Licenses. American Job Centers, Workforce Boards, increasingly we've been asked to use API to call the whole data set – they're being changed to get things ingested in bulk every couple of weeks.

Youth and young adult service providers, professional associations, wages by occupation and local area, job search, find related occupations based on KSAs.

Future –

COS learned a lot from previous meetings and Gary's work, so now they're being more thoughtful about how they're designed to make them more broadly useful. Naming conventions have been established and have helped working with developers – speeding up certain steps and improving communications. Now COS is spending a lot of time up front thinking about not just COS application, but also other use cases.

COS is building new APIs and improving old ones.

Federal data API work is happening and in the future this is going to be a core part of what we do, and the planning and collaboration with federal partners is critical.

Career videos – where are we with Spanish? Are there release dates added? Done with occupational videos. Support Spanish captions, no release dates. On Friday COS released industry and career cluster videos (about 12 – high level) – but still pretty useful as an overview.

Other interesting things people are doing:

Gary – As part of getting UI data to partner organizations quickly in the spring, Oregon delivered the code for an entire data visualization application via web service. This meant that the other department did not need developers to use the data and could very quickly integrate it in a website. There are potential uses beyond structured data – entire applications can be shared.

Matt – The ProjectionsSuite report manager now has an option for using a statistical model to get related knowledge or skill for occupations. Before a static threshold was used as a cutoff, but it resulted in some occupations having many related skills and others with none. The new process does a better job of identifying the relevant skills and is much more useful. This is available in the current version of ProjectionsSuite Report Manager.

NLX

How can we use job posting data as a supplement to LMI but without using a black box expensive service like Burning Glass. What can be done to enhance this data?

The National Labor Exchange (NLX) is sponsored by NASWA partnership – member organization, many large companies use them for OFCC compliance reporting. Other thing is that they collect job openings every night from every state job bank.

Around April we started getting the actual files which has made huge improvements in the use of the data and allows a comparison over the year.

COS has been using NLX since 2016 to create "hot certifications". They tried some products, but none worked that well. This has been improved a few times and is very popular.

2019 – NASWA wins 1 million dollar NSF planning grant to set up analytics system with NLX data. How could we set this up to analyze? Interviewed states? Just had an online expo highlighting these projects. Plans for the future – we're a pilot state to look at data. James developed some reporting for DEED leadership. Started getting the whole file in April and no longer have to request 500 at a time.

Having one person doing this work is not feasible long-term, so we made a collaborative SQL framework. NASWA applied for a second round grant and in it, they asked for partners, and we've volunteered – have a written proposal and what kind of products we could get.

How many other states are partnering? VA started in first phase. There are concerns about whether they can create something that works for all states.

The other interesting work was around using machine learning to improve job seeker filtering in the job search. Built programs that look through each record and identify highest level of education, licensed, filtering for jobs that have a veterans preference (not many), mention certification (slow), once you've obtained an ONET or SOC for the posting. Office, clinical setting, accurately pull out mention of pay. Most jobs don't bring that up. For 10-20 percent of postings there is some kind of specific wage data.

The regional detail is unparalleled.

We're extracting values from the posting data using sophisticated parsing algorithms, not AI. Microsoft will work with people who have complicated regular expression processes to make AI process to train it to know more types of things and that's being considered.

Commercial job banks aren't great about returning appropriate jobs – anything that can be done to improve the search and matching would be better.

When we get further along we might call another meeting for reporting, etc. The questions we'd like to address are how this can be used to enhance or add to the suite of LMI data – JOLTS at the state level, change indexes. How representative of the economy are online postings? What kind of strong alignments are there? Where does reality deviate from online job postings and how do we correct for that?

_____

ARC Meeting 1/14/2021

In attendance:
Patrick Flaherty, Rajani Edupalli, Dana Plazcek, Michael Peery, Barbara Ledvina, Nicole Kennedy, Christopher Robison, Kevin Doyle, Steve Duthie, Matt Steadman, Gary Sincick, Steve Hine, VenuGopal Japa, Steve Williams, Al Sylvestre, Bill McMahon, Mike Sylvester

John Pearce is officially retired, but is around part time in the short term
Steve Hine will be retiring at the end of the summer-Sept 1 (2021)
Barb Ledvina will be retiring April 2, 2021

LEWIS Update

Matt Steadman presented: All obstacles and bureaucracy preventing LEWIS from using its own Gov Cloud account have been removed and all are set up and ready for use. The team is now trying to formalize access accounts.

Steve Hine asked if that access will state cloud.gov allow access by non-Utah states once we're ready to go live?
The AWS access is for working on the program, but all the uploading/downloading will be handled by application itself.

Annually users have to validate their personal email address and an email address that ends in bls.gov as a means of ascertaining confidentiality training. They considered other actions and asked about a means of getting access to the list of BLS agent agreements, but BLS has no APIs or anything they could use and they confirmed this was an appropriate alternative method of confirming legitimate access.

Users will have access to their own state's data and cross states' data where there are relevant MSA areas. BLS staff have their own level of access for review, including to logs to see who is using the application.

There was no demo for this meeting. While they did get some things working on the test environment, in the new production environment they haven't yet gotten it ready for presentation.

Request for feedback
ProjectionsSuite will be facing the same transition to the cloud soon and since both are developed by the same team and serve the same LMI offices, Utah would like them to be partially integrated for the convenience of users. For the public-facing aspects of both sites they will be using Drupal. They plan on a single source of user authentication for both sites, allowing users who have permissions on both to sign on only once and access both sets of features. How tightly coupled should the two be? It's more user friendly if they're more closely coupled, but the two programs have different funding streams.

Steve Hine asked if Projections necessarily requires sensitive data access and corresponding user authentication. Nicole Kennedy noted that they do get non-public extracts compiled in NC. Kevin Doyle said it's minimal access, but all are BLS agents.
Gary Sincick suggested that decoupled might allow integrating new changes down the road.

There is a necessary degree of communication between the two – enhancing the ability to import data to the Projections side is an opportunity. Staffing patterns will be importable.

Christopher Robison noted that in NV very few Projections users have LEWIS access and OES is protective of that access – will that division be possible to maintain? Each side will have a group leader with the power to boot users. There will be some kind of final checkmark approval by OES group leader for Projections to receive requested data. "Published final data" is a permanent designation that affects suppressions and will be the only kind of data that can be exported. This will be more a push from LEWIS than a pull from Projections.

It was asked if PMP should be the primary driver of this decision, but that project is in an earlier phase. While the website will cut over soon (just waiting on ETA verbiage decision), the application is still in a planning phase.

Conclusion:
In most ARC states, there's not that much overlap between Projections and OES staff reducing the convenience factor. The different funding streams and security needs are a concern. We'd favor a less-tightly integrated solution but PMP's input is critical.

Follow-up:
This should be taken to PMP. Kevin Doyle and Christopher Robison are both on PMP committees. Mike Sylvester noted that there's a PMP meeting in February, likely the 18th. He will ask if a small group of ARC people can attend.

Employer Database

We've been asked to evaluate and potentially promote some DataAxle APIs. Historically that wasn't part of our contract and distribution has been through a download site and CDs. Bill McMahon noted that the last time this came up was with Jim Winner 2-3 years ago. InfoGroup was selling them as an add-on service and for LMI use there were a few problems: 1) The API structure is totally different than the WID structure that's shared with us. 2) the functionality was really limited – 1 or 2 records at a time.

They wouldn't work for production website and were not sensible for a data delivery mechanism given our analytical goals.

DataAxle has revised and expanded their APIs and want to offer more. It's unclear if access is a part of our new contract or if they are again selling us additional services or trying to replace the download site. ARC use of the data as a bulk analysis tool instead of a phonebook is fairly unique and it's unlikely their off-the-shelf solutions will be tailored to our needs.

CareerOneStop already has an API used to power their own application that enforces our user/publication rules and can offer access to that to states if they want them. Outreaching APIs could be made a priorty.

Steve Hine noted that we used to have intermediate user license agreements. All had an agreement with InfoUsa, but now we don't have that. How will DataAxle monitor access to theirAPIs without formal agreements in place for legitimate users? COS can't monitor terms of service, even though they can tell who is accessing their APIs. There's an amount of faith in the willingness of users to adhere to agreement.

Gary Sincick noted that there's an ongoing problem of hackers trying to pull more data. They block IPs when problems are spotted, but sometimes it takes time. If a state application is using an external API it looks to the provider like the state is abusing the rules – does this lock out the legitimate applications? How do you manage that? This is an issue for everyone. Bill McMahon noted that for this more sensitive application we'd have a very limited number of state users rather than all potential and the relationship would be more like partners.

Follow-up:
Steve, Bill, Gary, and Amanda will arrange a conversation with Don to make sure he understands the non-technical implications of this hazy "distribute via API" concept.

Steve Hine will work on the standards document for the new contract.

WID 3.0

WID 3.0 structure has been discussed. It involves breaking changes (name changes, significant changes to core lookup tables) and would require a longer comment period than usual to implement.

Steve Hine queried if this list of changes worthy of a 3.0 designation. The explanation was that the name changes have much bigger implications for dependent applications and yes, you do have to call it a major change and make it 3.0. The planned geography changes (to accommodate the more frequent MSA definition changes) is the real factor worthy of 3.0. It would also make sense that it would facilitate API project because we then wouldn't build the standard on the old structure and obscure field names. There's tension between general philosophy of platform agnosticism and things that are just more common now. Even if we're planning ahead for API definitions, it's not necessary that APIs will need to be ready to launch at the same time. The 3.0 release will require more technical support than a minor release and some proposed supports would be the creation of views that would replicate old 2.X table structure.

Amanda Rohrer questioned whether we offer enough benefit to states to require changes that have concrete costs beyond their usual level of support; while maintenance is funded, the grant hasn't increased in recent years and not every recipient has direct control over implementing structure changes. Dana Plazcek noted that that's a common argument against change and has resulted in this situation where we're using FoxPro naming conventions even though no current software is so limited. Many states also use a contractor like MT or GeoSol to host the database and aren't directly responsible for implementation. There may also be options for obtaining additional grant funds, but that can't be explored until the feedback process starts.

Amanda Rohrer noted that the feedback we get from other states can be inconsistent – our contact lists are opt-in and sometimes outdated. The people who sign up may have no responsibility for implementing structure changes or may be so busy they don't get the notice soliciting comment. Bill McMahon indicated that ETA funds contact list verification and if we have a starting point, we can ensure it's kept up to date using that method. Dana Plazcek also noted that Don Haughton mentioned another list at the last meeting and we should look at that as a means of publicizing the change.

Conclusion:
We will proceed with the feedback process.

Follow-up:
Dana Plazcek will prepare the documentation to release for comment.

Part of the documentation should include breaking out the changes by the reason – primary drivers, cleanup, and ancillary things we're fixing because we have the opportunity. This may come later in the process.

Amanda Rohrer will begin a contact list that is more specific to possible stakeholders – management, technical, etc. Once started, we'll try to fill in some of the gaps through known contacts on other committees or ETA program officers and keep it maintained through the COS process. This will enable us to announce the comment period in a more targeted way.

Licenses Admin tool

Amanda Rohrer talked about the license submittal process. While we've made significant improvements in data quality in the past few years, further progress is inhibited by the exchange and maintenance of data. We've got two types of state contacts – those who are technical staff, comfortable with transferring files and the structure but not typically involved in the creation of data, and those who are staff on other BLS programs like OES who may be more comfortable seeking out licenses and assigning codes but less comfortable with the submission format itself. The current process gets bogged down in the structure and the intersection of those two skill sets. It also is hampered by the speed of feedback – states submit files and may wait several days until it's loaded and evaluated, by which time they've moved on to other projects and are less responsive to concerns.

Since we're not currently covering the costs of meeting in person, there are funds in the budget. Building an application could solve some of these problems – it would give more technical submitters a bulk import option and immediate feedback on missing licenses or potential errors and would allow less technical users to page through licenses for review in a more approachable way, again giving immediate

feedback on problems. This would reduce frustration by submitters and help states manage the responsibility for the task with less training burden.

The Department of Labor has made good use of previous improvements and is actively pursuing more data and options for credentials, particularly those that aren't associated with a degree. There's good reason to believe that this will continue to be a priority.

Follow-up:
Amanda Rohrer has written a high-level description of the process. COS and Utah will take a look at top-level requirements.

CareerOneStop

COS has pushed out a new API of occupational keywords to codes with a choice of code matches (SOC, ONET, SOC-OES). That and other APIs will be converted to SOC 2018 January 29th.

National ETPL data. There is a new site – trainingproviderresults.gov – that includes ETPL data for all states and is searchable with the number of participants and graduates of the programs. COS has been approached to take on a distribution role, and this may involve the ARC for promotion or reviewing download files.

Data analytics – there is a dashboard group at ETA looking at COS APIs and Burning Glass data. We're looking at NLX data as an alternative. Amanda Rohrer showed some top-level comparisons of the various job posting data sources that have been a part of that discussion. This is in early draft stages.

All available sources:
https://public.tableau.com/views/AllSourcesComparison/ComparingJobPostingTrendsfromAvailableSources?:language=en&:display_count=y&:origin=viz_share_link

More detailed comparison between our NLX work and Burning Glass public files:
https://public.tableau.com/views/NLX-BurningGlassComparison/Overview?:language=en&:display_count=y&:origin=viz_share_link

Next Meeting

We will meet remotely next in April 2021.

Appendix B License Admin Tool

State License Review Process

Currently states send files to us, and I reconcile them with past versions and incorporate changes. There are a few problems with this setup:

1) The transfer of files is clunky. Many states have restrictions on attachments to email, the import tool I added to our website has pretty strict limits in type and size, and most states don't have dropbox or other. This confuses and frustrates state users.
2) Many times the person assigned to review the data has no particular knowledge or expertise on occupational licenses or data structure. They struggle with the requirements themselves.

Having a review process that walks them through it step by step would reduce the burden on states to develop a process.

3) If there are missing licenses or junk introduced, they don't always respond to questions. A more immediate review process would allow them to get feedback in real time and improve understanding of the issue.

Users

Each state should only be editing their own data, so users should be approved to start with. States can have multiple users and changes should be tracked by who submitted them. Password reset once a user is established should be automatic.

Loader

There needs to be a way for states that do have an existing process to submit standard WID-format files. This should accept multiple file types (Excel, txt, csv, Access). It needs to have a validation process (check state FIPS code to ensure user has access to state, text/numeric types). A file share option that's not limited by structure would be helpful as well.

Manual input

For some states, understanding the table structure itself is burdensome. Having a process where each license and its details are displayed on a screen as a detail page rather than a record and with codes displayed in descriptive format so that the user can flip through them and manually edit, add, or delete would be a significant improvement for many states.

Edits

There are a lot of things I check when I get a new version – that codes are the same so that values derived from other sources still line up, the total number of licenses, the presence or absence of commonly licensed occupations to see if they're missing. I can write up a process for checks that the user would have to sign off on regardless of whether manual or bulk entry.

Reconciling

Currently there are about 5 different streams of data being brought together. There are some other sources for licenses, plus licenses I've collected, plus the state submitted data. Sometimes a whole license or license authority is brought in, sometimes it's just the details that are added (some fields but not all). This is already established in the process, but it needs to be incorporated in the tool and the source displayed so that users can understand where a value they didn't input came from. I see this mostly as a need for labeling data elements in the edits or display.

Occupational Coding

This process is largely automated already, but it's designed to be run against the whole set of state licenses rather than an individual license. There are two options – the process can be run in between the import/revise stage and the edit stage against the whole set, or a simplified process could recommend possible occupational codes to the user. In general, I only adjust the occupational codes when licenseids have changed or are new, so the second option of recommending codes may be more appropriate.

Approval Process

Even with an edits process, there are likely to be some changes that aren't what I want them to do. There should be a step where I can review their finalized data. Prior vintages also should be preserved in case of errors.

Export

States need to be able to get their data out either for review or other uses. This can be WID format (same as input), but should have a few options (csv, Excel).

Alerts

Users (and maybe supervisor?) should be alerted when new data is due or if edits process hasn't be completed.

Appendix C State Published OES regions

stfips	areaname	OES region type
01	Alabama	state, msa, bos
02	Alaska	state
04	Arizona	County and WIOA
05	Arkansas	wda, state, msa
06	California	county,state,msa
08	Colorado	state, msa, bos
09	Connecticut	Labor Market Areas, wda
10	Delaware	state
11	District of Columbia	dc
12	Florida	state, msa
13	Georgia	WDA, service delivery area, county, state, msa
15	Hawaii	MSA/County
16	Idaho	Labor Market Regions
17	Illinois	Economic Development Region, LWIA
18	Indiana	msa/bos
19	Iowa	msa/bos
20	Kansas	Projection Regions
21	Kentucky	LWA
22	Louisiana	RLMA
23	Maine	state, county, msa, balance of state, wia
24	Maryland	state, msa, WIA
25	Massachusetts	state, wda, nectas, 9 divisions
26	Michigan	state, msa, bos
27	Minnesota	Economic Development Region
28	Mississippi	state, MSA, WDA
29	Missouri	state, regions, msa
30	Montana	state, msa, oes regions

31	Nebraska	state, msa, metro/non-metro, wia, economic development region
32	Nevada	state, msa, bos
33	New Hampshire	state, county, msa
34	New Jersey	state and msa, wib in the future?
35	New Mexico	state, msa, workforce regions
36	New York	state, msa, labor market regions
37	North Carolina	Prosperity Zones
38	North Dakota	county, msa, state
39	Ohio	state, msa, jobs ohio regions
40	Oklahoma	state, msa, bos
41	Oregon	state, regions, counties
42	Pennsylvania	state, county, msa, wda
44	Rhode Island	state
45	South Carolina	county, wda, msa, state
46	South Dakota	state, msa, bos
47	Tennessee	state, msa, bos, region, wda, rural, historical regions
48	Texas	state, wda, metro
49	Utah	regions
50	Vermont	necta and two bos areas
51	Virginia	planning region, workforce development region, community college region, state, bos
53	Washington	state, msa, bos
54	West Virginia	state, wia
55	Wisconsin	state, wda, lsa, other
56	Wyoming	regions
78	Puerto Rico	puerto rico

Appendix D Future Investigations into Real-time analytics options

NLX Future Projects

Standardizing methods

There are a number of questions people may ask of the data that could be answered in different ways depending on the application or the assumptions of the analyst. Best practices or common definitions for what those are should be developed and documented. In many cases we've developed these methods independently, we just need to reconcile them between analysts and consider if they're universal or just convenient.

- Prioritizing occupational codes

  The autocoder returns multiple matches for every posting. In the raw data those have no indicator of the quality of match. Because of the size of the source data, we don't re-run autocoding on every job, only those with bad matches. As a result, the sum of occupational

totals will be greater than an aggregate that doesn't account for occupation. When to use the random/first match code and when to use all codes should be defined.

- Time reference point

When attempting to tally postings for reference time periods like days, months, or years, there are several approaches. A reference day within the time could be used. New postings opened during the span could be used. An average of those open on any day during the time span could be used, etc. To ensure meaningful and consistent comparison those should probably be calculated in advance and stored in lookup tables.

- Weighting jobs

There is an optional field for the number of open positions a posting represents, but its contents are unconstrained. Besides complete junk (non-numeric characters, punctuation, etc) there are some ambiguous values. "A01" or "9999", for example, have to be interpreted as either numeric, error, or placeholder values and treated accordingly. Most importantly, that has to be done early enough in the process that all analysts use the same rules.

- Reconciling submitted and supplemental variables

In some places there are infrequently populated user-submitted structured data fields that duplicate a similar concept to what we're pulling out of the text description of the posting and that don't necessarily agree in content. Which value is used or prioritized should be consistent across users.

- Geography

The data contains several geographic variables, but for meaningful comparison to government data we would need to assign statistical regions, such as counties or MSAs. This is fairly simple, but could be done from address or zip code or city. Null values need to be handled, as well. Depending on choices made, output could get different results.

- Employers

While we get a company name and posting url associated with each advertisement (the root website could be a sign of a common parent company), that's not an identifier. FEIN is included for some but not all postings and we don't know how accurate it is. As a result any analysis that tries to identify distinct employers or particular employers will have to account for situations where the punctuation in the name is slightly different or there's a suffix like ", Inc." or "#123" following a parent company name. I've already partially written a process for standardizing some of that, but that's another cleanup task that should happen at the front end so it can be applied consistently.

Standardizing interpretation

Given the nature of the data, there's as much absent from it as present, and the way that's interpreted and explained should be consistent. In some cases, further research is needed to identify exactly where those gaps are.

- Geographic detail

   The data contains geographic data down to the zip code. While superficial checks suggest that MSA/non-MSA breakouts are roughly similar to state proportions, a more systematic review should be undertaken. This would cover relative representation of different states, how granular we can get and still trust the geographic data, and if there are particular employers or industries that should be treated differently because of their hiring practices – long-haul trucking or employment services, on the industry front, or large employers with certain hiring functions handled at regional offices on the employer level.

- Coverage by state and occupation

   Some big-picture and more qualitative review should be considered at the state level. If we're going to use this to compare state hiring trends, we should know if the universe it represents is measurably different between states. One question would be the number of included employers or the policies of state job banks that are contributing to the data. Differences in seasonal or occupational patterns resulting from these policy differences should be noted.

- Changing universe

   The employers submitting data to the NLX are changing all the time, largely increasing in number and variety. To do any kind of time series analysis, that would need to be tracked or adjusted for.

- Seasonal hiring

   As we develop more of a time series, one of the areas that is interesting is identifying when given occupations increase hiring and what qualifies as a boom.

Value to be added

There are variables and ways of looking at the data that aren't already collected that could be added as a secondary process.

- Industry

   While the classifications that come from the source may include industry, the majority of records don't have it associated. This could be added for many either by comparing the standardized company name to other sources of data or by imputing it based on the occupation or some combination of the two. While this would be fuzzy with available and public data sources, it could be meaningful at the 2-digit level.

- Firm size

  A lot of analyses would be improved by knowing firm size. This would likely be obtained as part of the industry process above.

- Emerging occupations/job titles

  One of the real strengths of data this current and detailed is what it can tell us about jobs that are either more specific or too new to have a home in the SOC structure. There's a lot of potential for monitoring the frequency of specific keywords in job titles, such as "Zoom Coordinator" or "Contact tracer"

- Any other supplemental variables?

  In initial analysis, comparing to JOLTS data was somewhat stymied by the lack of a start date variable. It matters to their definitions, but we don't check for it. That could be added either in the python process step or later on.