

National BILT Meeting Minutes
“Data Analytics” KSA Vote and Discussion

MEETING DATE: Tuesday, November 19, 2024	MEETING TIME: 9:30am-12noon Eastern	MEETING PLACE: Zoom
RECORDER: Mark Dempsey	RECORDING: Available upon request	PREVIOUS MEETING: MegaBILT trends meeting – August 27, 2024

MEMBERS PRESENT

BILT:		
Jeffhram Balilla	Bob Hitchins	Priti Malkan
Jon Brignac, USME	Dan Huff	Ginny Proctor, Ohio Dept of Jon and Family Services
John Chioffe, Amtrak	Corey Kirkendoll, 5K Technical Services	Ravi Tandon, HCL Tech
Craig Cocciola, Cogent Cyber Range	Jeff Magnuson, SpartanNash	Plus 2 more employers
SCHOOLS:		
Calhoun Community College	Houston Community College	North Arkansas College
Chandler-Gilbert Community College	Idaho State University	North Central Texas College
Columbus State Community College	Jacksonville University	Roger Williams University
Forsyth Technical Community College	Lansing Community College	Sacred Heart University
Fox Valley Technical College	Lone Star College	Trident Technical College
Grand Rapids Community College	Luzerne County Community College	
NITIC staff: Ann Beheler, Mark Dempsey, Christina Titus, Stephanie Schuler, Larry McWherter, Leah Palmer, Diane Meza		

Agenda items	Discussion
Introductions and KSA overview	<p>Larry welcomed attendees and explained he is the principal investigator (PI) for the National IT Innovation Center grant.</p> <p>He discussed NITIC, which is now in the second year of a five-year grant from the National Science Foundation. NITIC’s goal is to help schools get graduates ready to work by helping keep curriculum current with industry needs as provided by BILT meetings like this. Larry briefly explained NITIC’s initiatives: offers education resources through an online clearinghouse, promoting the implementation of the BILT model, hosting an IT Innovation Network community of practice of IT educators (100+ colleges across 32 states as of now), hosting faculty professional development events (which will be transitioning to a “year round” model), and developing models to increase diversity and include all student populations.</p>

	<p>Ann noted how much NITIC has already accomplished in the first year. She explained that the goal of today’s meeting is to update entry-level job skills for data analytics. The last time this list was updated was 18 months ago through the IT Skill Standards project grant, which has since sunset. Consensus is not necessary. Employer SMEs are the only ones who will vote on the job skills. Educators are at the meeting to listen and answer questions as needed. Ann also noted that NITIC will host a webinar series in the spring to help ITIN educators implement BILTs at their school to better align curriculum with industry needs.</p> <p>The final prioritized job skills will be shared with educators across the country so they can “localize” it with their own group of employers. Dissemination of the list will happen via NITIC’s website, the ITIN community, national conferences, and LinkedIn. In addition to the prioritized job skills list, NITIC will share anonymized meeting minutes. Ann understands some SMEs are volunteering their time unofficially.</p> <p>Ann explained that NITIC’s grant team, ITIN educators, and employer SMEs are all on the call. Rather than go around the room for introductions, Ann shared a slide with the names of those SMEs who planned to attend. Ann told the SMEs that they’re expected to fully participate in both the vote and the discussion, but need not share anything proprietary. Ann will facilitate the discussion and make sure everyone participate, even those who are quiet.</p> <p>Ann explained the 1-4 voting scale. Each job skill item will receive its own individual vote.</p> <ul style="list-style-type: none"> * “4” means absolutely must be in curriculum * “3” means should be in curriculum * “2” means it’s a “nice to have” * “1” essentially means delete <p>The “1” votes are important because we need to know what to take out to make room for new items.</p> <p>After discussing the vote results, Ann will ask the group to suggest items not on the list that should have been there.</p> <p>Mark noted that “K” knowledge items refer to “awareness.” At a previous meeting, there was confusion that “K” meant something someone could do. This is incorrect; it’s the “S” skills that refer to something a person can do.</p> <p>Using a QR code (and URL in the chatbox), the SMEs voted electronically.</p>
Knowledge	<p>Ann briefly explained that items getting an average vote of 2.6 or less often ends up getting removed from a program. Prioritizing knowledge areas and skills are important because items that are not important can make room in curriculum for more urgently needed items. Ann noted that BILT members will have the chance to advocate for keeping any item. with a 2.6 average or less. BILT members can also modify the wording of an item.</p> <p><u><i>K1 - Knowledge of risk management processes as part of a software development cycle, 2.63 average</i></u></p> <p>One SME voted this a 3 because in their environment risk management is very important “across the board.” If you’re developing apps or even using apps to gather and analyze data you need to be knowledgeable about the risks. You need to know the controls associated</p>

with the app or data set you're using. Another SME agreed – he voted a 4. He's seen a lot of entry-level employees start the job with the ability to code and understand data, but few of them "have any concept" of what risk means to an organization. They don't know the basic fundamentals. Another SME (works for a large company) also voted it high – they find this lacking in new hire. Still another SME (consulting for large to mid-size companies) agreed. All employees need to have a basic understanding of this. Another SME noted that while the software development cycle isn't a core part of analytics, it does play a role in predictive analytics. Another noted that lot of analytics can be done with tools like Tableau without ever going into coding.

K12 – Knowledge of command line tools (e.g., mkdir, mv, ls, passwd, grep), 2.25 average

Ann asked the BILT is anyone wants to lobby to keep this item. One SME stated that this isn't essential for an entry-level hire. Even if someone is using Python for modeling, they'd likely be using existing models to do the work. Most employers would not give command shell access to new employees. Knowledge here may be good enough. Would the new hires ever actually use the commands? This SME didn't think so, unless the employee was working on the infrastructure side. Ann wondered about covering this at a high level in the classroom. That SME agreed – for him, it was the tools that concerned him. It's good to have a knowledge of the tools like PowerShell, but not the individual commands. He added that he'd rather suggest to students that if you have to use Command Line shortcuts to get the data quickly, it's better to use Jupyter notebook trackers rather than the shared command lines. None of the BILT seemed to believe this was essential as a K or an S.

ACTION: consider removing K12 from the list, but adding PowerScript as an "e.g." example.

Ann asked employers not to be distracted by the "such as" examples – IT faculty will use whatever "contemporary products" are available at the time.

K13 – Knowledge of interpreted and compiled computer languages, 2.44 average

One SME voted 2 here because entry-level workers are not building models or programming. Python is "in the mix" but Python can do both. Another SME (large company) agreed: basic information about interpreters and compilers is good but entry-level workers won't be using them.

ACTION: consider removing K13 from the list

K14 – Knowledge of how to utilize current popular frameworks and languages, 3.00 average

One SME noted that it seemed redundant to include this K item in addition to other items that call out specific languages like Python.

K31 – Knowledge of ethics as it applies to data analytics and how to apply ethical judgment when policies are not well defined, 3.06 average

One SME found this item to be broadly worded. Ann noted that more and more across the country there is an interest in making sure students understand how to handle sensitive data sets ethically. Another SME agreed. He noted that it's the combination that could be risky. That is, two data sets are safe to work with separately, but combining them together makes them more sensitive to compromise. Sound judgment is essential. Another SME said that when he considered K31 to be about categorization. What is the category of data you're working with and what is the risk associated with that category? Another wondered about narrowing the scope to things like HIPAA so it's more about restrictions to the data more

	<p>than analytics. Ann explained that those kinds of specifics on laws and rules had been on the list in the past, but ultimately the BILT decided to remove it because that level of specificity was too much for a two-year program. Ann asked today’s BILT if students should learn specific standards like GDPR. One SME answered no – exposure to the standards is “too far” because they can be so specific and regional. Better instead for students to understand ethics and how to make good ethical choices. What are you protecting and what are the thought processes needed to do that?</p> <p><u><i>K33 – Knowledge of the landscape of BI tools (Power BI, Google) and data preparation tools and understanding of the data platform associated with each, 3.38 average</i></u></p> <p>One SME wanted to discuss the lower votes on this item. He wondered what the issue was since these are “next gen” visualization tools beyond Excel. They’re not really analysis tools. One SME responded that students should have knowledge of the tools, but as an entry-level worker they don’t need to understand the data platform. Ann acknowledged that K items are “awareness only” for students. Another agreed that digging into the underlying technology isn’t necessary, but students should know the tools. Ann wondered about removing the “data platform” statement. BILT wanted to add Tableau to the list of examples and add “such as.” Another SME agreed that PowerBI and Tableau are important tools in his business. In the chat box: Now that python is available in Excel that might change the conversation.</p> <p>ACTION: consider adding Tableau to the “e.g.” example list in K33.</p>
Skills	<p><u><i>S6 – Skill in writing code in a currently supported programming language (e.g., Python), 3.31 average</i></u></p> <p>There was discussion about including R as an example. Ann noted it’d been included before, but then taken out by a subsequent BILT vote. One SME (government) considered R to be a “currently supported programming language” but admitted that could change. When she started her SQL training, they were 3 or 4 revisions back from what she was using in industry. Ann noted that the use of R seems to be regional. One educator (Arkansas) said their employers use R and Python. He joked that it seems to be a “religious issue” so they teach both. Another SME agreed that if students don’t have exposure to R, they will be “disadvantaged.” Another also agreed: both Python and R have pros and cons and depends on which data asset you are using. Ann suggested adding R into the examples with the understand that faculty will work with their local employers to determine how to handle it.</p> <p>ACTION: consider adding R to the “e.g.” list of examples in S6.</p> <p><u><i>S19 – Skill in preparing and presenting briefings, 2.88 average</i></u></p> <p>Ann asked about S19. Do employers want entry-level workers doing briefing? One SME didn’t expect new hires to do a lot of briefing. More likely, they’re handing their findings up to someone else. That said, this SME wants someone who can deliver findings in an understandable way so that he can go to the briefing prepared, rather than getting just a page of data. Students needs to know what goes into getting ready for a presentation. Another SME agreed, but suggested changing the wording to “skill in preparing for briefings and building presentations.” Another SME admitted his company “PowerPoint ourselves to death” but business communication – they can have PowerPoints of 100 pages – is essential.</p>

	<p>ACTION: consider rewording S19 to “Skill in preparing for briefings and building presentations.”</p> <p><u>S21 – Skill in using multiple search engines (e.g., Google, Yahoo, LexisNexis, DataStar) and tools such as ChatGPT in conducting open-source searches, 3.06 average</u></p> <p>One SME asked about the context for using a search engine in data analytics. He noted also that no organization will allow a direct open connection to ChatGPT. Another SME responded that from an analytics perspective, anyone coming into the workforce now ought to be able to search Google and find answers on their own. This includes knowing when the answers search returns may not be correct. It’s not just about the validity of the answer, but also the validity of the data source. The BILT suggested rewording the skill to “skill in validating the data that comes up after using search engines.” Another SME (Fortune 500) read S21 as a kind of basic prompt engineering. How you phrase the prompt or the search will impact the results. New hires need to effectively ask for the right information. Data analytic curriculum should include prompt engineering.</p> <p>ACTION: consider rewording S21 to “Skill in validating the data that comes up after using search engines (e.g., Google, Yahoo, LexisNexis, DataStar).”</p> <p>One SME wondered if this is a duplication of S19, but the BLT decided these are different skills. Presenting is a different skill than telling a story.</p> <p><u>S30 – Skill in identifying basic common coding flaws at a high level, 2.63 average</u></p> <p>One SME scored this low because S30 finding common coding flaws is so “language dependent.” I scored S30 low because it’s very language dependent. This is not a useful skill at this level. Ann wondered about making it read “dependent on the language you learned” but the BILT didn’t find that applicable. This may be better as a general knowledge item, rather than skill in identifying. Another SME added that as data analytic tools mature, there’s less of a need for coding. What used to be a manual process for analysis is now included as advanced functions inside tools.</p> <p>ACTION: move S30 to the K section; this is a knowledge area, not a skill.</p> <p><u>S36 – Skill in recognizing and assisting in mitigating cognitive biases which may affect analysis, 2.81 average</u></p> <p><u>S37 – Skill in recognizing and assisting in mitigating deception in reporting and analysis, 2.75 average</u></p> <p>One SME noted that mitigating biases is more on the AI side of the world. The tools will predict and analyze. The BILT voted these high but in the discussion, there was concern about the word “mitigate.” That’s not something an entry-level hire will do. That’s mostly done in the modeling.</p> <p>ACTION: remove the word “mitigating” from S36 and S37.</p>
<p>What to include next time</p>	<p>Ann next opened up the floor for the BILT to propose what items should be included on the next vote.</p> <p>Prompt engineering is important. Ann noted that prompt engineering wasn’t even a term 18 months ago. One SME noted that generative AI goes alongside prompt engineering.</p>

Anywhere on the job skills list where we list ChatGPT, “gen AI” should be listed as well. While one SME stated that successfully using a search engine is a kind of prompt engineering, another SME disagreed – while search involves keywords, prompt engineering more of a dialogue. Others asked if prompt engineering really belongs in a data analytics program. Isn’t prompt engineering more aligned with AI? Maybe “knowledge of prompt engineering” is better. Another SME countered that generative AI tools – including ChatGPT – can be used to produce analytics. Her suggestion is for students to learn how to use gen AI, understand how it works, and how to use prompt engineering to build the database you want. Another SME added the term “agentic AI.” Generative AI generates content, agentic AI takes action on that content. Agentic AI will be coming soon.

One SME asked to include Excel alongside examples of PowerBI and Tableau.

One SME asked about Databricks, which can be used with both analytics and engineering. He’d like to vote on that next time. Another SME agreed students need to know about platforms like Databricks and Snowflake. But another SME asked if they’re talking about ETL (extract, transform, load) knowledge or something else. The reply: not the ETL portion, which is engineering, but more about doing a kind of data analysis, making sense out of the data using Databricks or a Jupyter Notebook. Another SME noted that all of the platforms now have their own AI features (Databricks has Genie). So while Databricks is a data platform by default, it’s enriched by analytic tools.

Note also that Databricks came up in the S11 discussion as a tool example to maybe be added to the Excel and Python examples.

Databricks: <https://azure.microsoft.com/en-us/products/databricks>

Snowflake: https://en.wikipedia.org/wiki/Snowflake_Inc.

Ann asked about cloud computing knowledge. Do students need to know that? One SME suggested new hires would just be using the tools. They likely won’t need to know if the tool is in the cloud or on-prem. Another SME wondered if the question is really about students knowing the “data stack” inside cloud providers AWS, Azure, and Google and how the data moves through them. The knowledge would be useful, but students don’t really need to know the technology around that. It’s more about understanding the data flow. These are knowledge items, not skills. Another SME noted that Databricks is available for all three providers. Another SME wondered if “data stack” is too broad a term since it could also encompass things like PaaS, SaaS, and IaaS. He suggests zeroing in on just the data offerings.

How prevalent are data warehouses? One SME agreed there is a difference between databases, datamarts, data warehouses, and data lakes. Students should probably know the difference. Another SME noted that this seems to be about data architectures awareness. There are many different services used across the platforms for data prep and storage.

Data lake versus data warehouse: <https://aws.amazon.com/compare/the-difference-between-a-data-warehouse-data-lake-and-data-mart/>

Ann ended the discussion with an explanation that employee behaviors (i.e., soft skills) are covered by the NITIC grant separate from technical content like today’s meeting.

She asked the SMEs to email the NITIC staff if think of anything later to add.

Next steps	<p>Larry discussed next steps.</p> <ul style="list-style-type: none">* The outcome of this meeting (prioritized spreadsheet, meeting minutes) will be shared before the December break.* Tuesday, January 21 will be a “megaBILT” meeting inviting SMEs across all IT disciplines to share their outlook on emerging industry trends. This is another way to help keep educators updated on workforce trends. <p>Sharing prioritized data Fall 2024 before the holiday break</p> <ul style="list-style-type: none">* In Spring 2025, the NITIC grant will convene a similar “vote and discuss” meeting as today, but focused on AI job skills. <p>All of NITIC’s BILT meeting results are posted on the grant’s BILT page: https://www.nitic.org/industry/national-bilt/bilt-overview/</p>