

## WHAT DATA AND DATA TOOLS ARE MOST NECESSARY TO MAKE POLICY AND PROGRAMMATIC DECISIONS?

**Dawn Thomas, Ricardo Covarrubias Carreño, & Keith Hollenkamp**  
*Illinois Early Childhood Asset Map, University of Illinois Urbana-Champaign (USA)*

### Abstract

What do early childhood administrators, policymakers, and advocates want in a data tool? It is easy to simply provide them with multiple datasets, geographic regions from which to choose, and lots of demographic data describing young children, their families, and communities. But is that what the early childhood community needs or wants? Working with other stakeholders across the state, this early childhood data hub interviewed key stakeholders, held focus groups, and investigated what the statewide community wanted in their data and data tools. As a result, existing data tools have been renovated and developed with the early childhood community in mind. This project considered the needs of stakeholders and users based on their data literacy. How does a data tool and its datasets bridge the divide that may exist between participants and users with differing capacity, interest, and knowledge of data? How do we build a foundation from which to develop higher capacity in using data, more confidence in using data tools, and better understanding of effective uses of data in reports, programming decisions, and changing policy? This presentation will provide an overview of the findings of this investigation and the tools that have been refurbished or developed based on those findings. The authors will explore the most necessary data, the quality of those data, and elements in a data tool that provide the user with the type and amount of data needed to make evidence-based decisions and answer specific research questions.

**Keywords:** *Data, data tools, policy, early childhood.*

---

### 1. Introduction

The state of Illinois in the United States is in the middle of a long-term data initiative, the Illinois Longitudinal Data System (ILDS), which involves redesigning public portals for ease of use and access of data, but more importantly, involves rethinking the value of discrete databases from each state agency to integrated datasets across all state agencies. This is occurring most predominantly in the arena of early childhood education and care data. The Early Childhood Participation Data Set (ECPDS) is in its fourth year of designing, building, and implementing integrated datasets from select standing databases from pertinent agencies. These agencies include the two responsible for most of the service delivery for young children and families—the State Board of Education (ISBE) and the Department of Human Services (IDHS). The Illinois Early Childhood Asset Map (IECAM) project was contracted to build and implement a new data tool, to be known as the Early Childhood Data Hub, and populate it with curated datasets from participating state agencies and data sources.

### 2. Rationale

IECAM (and partners') work has evolved from defining the process of the ECPDS to the necessary methodology to the reality of joining disparate systems' data into a single dataset using unique identifiers for each child to track their participation in different programs in different agencies. The work is done in a cloud-based environment. From the moment the curated data is uploaded into the cloud environment to the development of virtual standing datasets, from which to create integrated datasets across funding streams and programs, the data remains in the ownership of a particular agency. Researchers and evaluators from across the state who work within the confines of the cloud environment are instructed to develop data access and use agreements (DAUAs) with the data sources (i.e., agencies). Currently, the executive committee is working with the federal Head Start/Early Head Start program to ensure that its data is included in the initiative and meets the same governance criteria as other data sources and state agencies.

The conceptualization of the ECPDS project (1) began with the realization that many state agencies lack the necessary technology and technical know-how (i.e., capacity) to begin such an endeavor; (2) continued with the work of establishing communication and trust to discuss, much less share, data

between agencies; and (3) ended up in early 2024 having to meet challenges of the need for more funding, agency capacity, agency trust, and even more data with which to learn all we can about the children of Illinois. Pahlka (2023) insists that to make real change regarding the digital age and use of data, it is less about rushing to build new apps and tools and more about increasing governments’ capacity with data and technology. Although the culmination may seem to be developing the architecture of a data hub for users, collaborators across Illinois are learning the need to play the long game. The new Data Hub is being built on a durable foundation of trust, governance, collaboration, communication, and capacity building. We are learning it is less about the data that will be provided in the Data Hub and more about the ability and the infrastructure to look across datasets and answer questions that, on the surface, may seem easy, but are anything but. We are also realizing the expectations built into the development of the Data Hub for the state are themselves built on factors over which we have little to no control.

### 3. Methodology

Through a partnership with Northern Illinois University and the Center for Early Learning Funding Equity (CELFE), a third party was contracted to conduct interviews, focus groups, and a questionnaire of a specific group of stakeholders who either have used or will be using IECAM as a part of its current responsibilities. In September through November 2023, this contractor facilitated six focus groups to inform the design process of the new Data Hub on the IECAM website. The participants were shown draft designs for the Data Hub that would in the end offer enhanced functionality and flow of data. Participants were asked how useful they thought the Data Hub would be as well as suggestions for improvement.

The contractor sent an informal questionnaire to 181 contacts who were key stakeholders identified by the ILDS team and other respondents from an initial survey used in a previous project, and 58 responded.

### 4. Results

Participants responded with enthusiasm about the wireframe and the potential value of the Data Hub for presenting data. Although many of the comments were related to data questions, data quality, or availability, the results discussed here focus primarily on perceived usability of the draft Data Hub and suggestions for improving it (see Table 1).

Table 1.

Perceived Usability	Suggestions for Improving	Select Comments by Type
Potentially more user friendly than existing tools More accessible to “lay” audiences Ability to refer agencies and organizations to the tool Will save time and streamline processes for users Will serve multiple goals across different user types Will potentially answer specific early childhood questions unable to be answered at present. Useful for community needs assessments Useful for grant writing and reporting	Sets of “preset” views of the data: (enrollment vs. poverty, enrollments vs. demographics, enrollment vs. age) Contextual information for using the tool Native guidance for using the tool (e.g., orientation video and tooltips) Further beta testing to ensure its usability and accessibility	<b>Appearance</b> Visually engaging Visually overwhelming Colorful Too much text  <b>Data</b> Old Inaccurate Need definitions  <b>Functionality</b> User-friendly Sort of user-friendly Need tutorials

### 5. Discussion

While not surprised by the results, the contradictory nature of the comments was unexpected. For many of our users in the past and in the present study, their primary focus has been on their preferences (e.g., color, shape sizes, text fonts or sizes). Our users come from a variety of backgrounds and experiences with technology, digital tools, and data. What we have found is that in addition to capacity, knowledge of data, and ease with technology, preference or personal taste tends to play a large role in a user’s impression of a tool. One tool will not please every user. And, while data architects may choose a platform, a framework, and a language with which to develop a tool, for many users their preferences regarding color

or their proclivities for more or less white space on a screen may play an outsized role in whether they find a tool useful or not. This study aimed to answer three questions for IECAM as it developed the Early Childhood Data Hub:

1. What do early childhood administrators, policymakers, and advocates want from the Data Hub?
2. How could the Data Hub and its datasets serve participants and users with differing capacities, preferences, interests, and knowledge of data?
3. How can we help users and state agencies with differing levels of knowledge, ability, and technology use data tools more effectively?

### **5.1. What users want**

Results indicated that users want everything in one tool, yet they want it to be intuitive to every level of experience. They want it visually appealing, with pleasing colors and shapes, without, of course, being overwhelmed on the screen of their choice. They want functionality to be user-friendly, yet robust. Users want a tool able to deliver rapid results, with data that are current, available in multiple geographic regions (e.g., counties, municipalities, zip codes) and displayed in appealing, yet understandable ways.

Audience is important when developing a data tool; knowing our audience helps us better design a tool appropriate to our purpose. Having a clear view of our users enables us to acknowledge their capacity for using a tool, provide guidance in functionality, and have clear definitions of terms and data. Individuals in the focus groups concentrated on the intended audience and gave clear descriptions of what a tool should and should not include based on that audience. For instance, one user stated unequivocally that the draft Data Hub was too focused on the early childhood community and should be more open to additional age groupings. Another focus group member reported that they appreciated the fact that it was designed with the early childhood audience in mind.

Text can provide the boundaries of a tool, providing guidance and limitations of a tool and the data. Is the text overwhelming the data? Is the text complicating what could be a much easier tool to use? In other words, are we using too many words when the data and display of the data can speak for itself?

Users want the most recent data, but we are only able to provide the most recent data available. There is a distinction between what users want and what they might be able to get. (For example, demographic data from the U.S. Census for 2022 was released in December 2023.) Providing current data primarily depends on each data source. That means building relationships with the data sources, cultivating a rapport of trust with agency staff, and building capacity within the data source so they understand the richness and usefulness of their data, as well as the importance of disseminating and sharing their data. Even though available data may not be as current as users want does not mean it cannot be useful. “Old data” can still be used to look at trends and patterns, even when it is not the most recent. While our goal would always be to have the most current data, realistically, that may not be entirely possible. Ensuring that users are aware of the efforts taken to get desired data and the integrity to be honest when we admit we are unable to provide them with the most opportune data is crucial. Providing needed data is a transactional relationship and one that needs trust at its core.

### **5.2. Bridging the divide**

Acknowledging that our participants had high expectations for the Data Hub was critical in thinking about how we go about designing the architecture and building its functionalities. There seems to be at least three divergent approaches to handling differing perspectives and expectations for such a tool. The first is to make the decisions for the users. Tell them what they want. Sometimes people do not have the vocabulary or experience to know exactly what they desire in a new tool. This is completely natural and something we all experience. This option involves building a tool based off opinion (your own or others) and experiences (your own or others). An example of this type of approach is Steve Jobs and Apple. Of course, innovation is built on customer feedback and the reality of the market, but the original tactic was simply letting the users know what they wanted and needed in a tool. IECAM has never really taken that method. A second might be called the “more is more” approach. Incorporate as much as possible into a tool to see what works. Give more power to the user to interact with a tool. This does increase the amount of time and resources needed for development and may make a tool more difficult to use, but it provides users with everything they could need at their fingertips. (From IECAM’s experience, providing thorough definitions and documentation cuts down on redundant questions from users.)

This was IECAM’s initial approach. When IECAM was first developed in 2006, the original intent was to provide a set of interactive maps from which users could manipulate demographic parameters (e.g., poverty, maternal education, substance abuse) and overlay early childhood resources to show a full mapping of a particular geographic area in Illinois. Users could choose from a number of different parameters of demographics and ECEC programs and resources and produce their own maps. They could download, print, and use them in their own reports. While tabular data was offered, it was on a temporary

basis. But more wasn't more, it was mere muddle. What we found was that users were overwhelmed with choices and did not want to develop their own maps. The entire GIS (geographic information system) section went virtually unused for years. Slowly we began to simplify the maps, offering less parameters to choose from and, finally, we took all links to the GIS site off our main website. In addition to its widely used database that presents tabular data, recently IECAM has been publishing interactive maps with limited options and layers, and they are also quite popular on the website.

The third approach is to ask current and potential users what they want and need from a data tool. Keep those user suggestions in mind as you build your tool, but also rely heavily on your experience in building data tools as you sort through contradictory and unrealistic suggestions and expectations. Otherwise, you could end up with a confusing or contradictory range of choices or another "more is more" muddle. This is a balancing act that requires ongoing communication between the creators of a tool and its users, as well as the willingness to make necessary changes as time goes on. (Note that when it came to building the overall architecture, structure, language, or other necessary "back end" elements of the Data Hub, decisions were made by the IECAM team based on their extensive knowledge and expertise.)

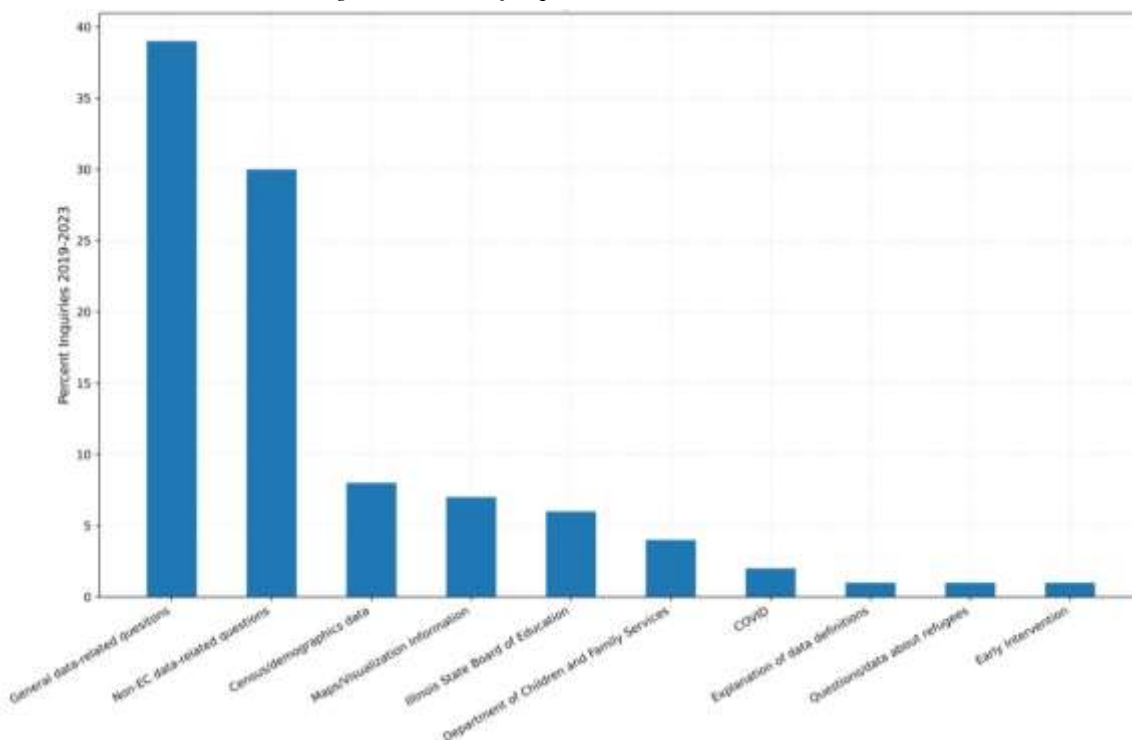
### 5.3. Developing capacity

IECAM is a publicly funded website with aggregate public data. Although users can find all the data on our website from other data sources, we offer a one-stop shop for early childhood and demographic data, descriptions and definitions of those data, charts, tables, and interactive maps and visualizations. While we have conducted focus groups in the past looking at functionality, appearance, and ease of use, having the chance to find out what current users want in a new tool before building it presented us with opportunities and challenges.

We identified a couple of issues that will need to be addressed and/or acknowledged moving forward. First, we cannot please every user. We can listen to users. We can respect their suggestions and thoughts. We can implement changes to some of these suggestions. However, we will not be able to build a tool, much less improve our existing website and tools, that will satisfy everyone.

Knowing that and being left with contradictory comments from the 2023 focus groups, we decided to look back at our own records. Users have been able to request data, charts, and custom maps. They can request assistance through telephone calls, emails, an electronic request form, and a chat box on the website. Users may ask for specific data types from a particular geographic region; they may ask for charts, tables, or maps to insert into their program reports. Some ask for information that requires staff to refer them to outside resources, websites, or organizations. IECAM has been tracking all data requests since 2008 and logging information, such as a description of the data request, geographical location of the request, organization of the requestor, and institution of requestor, if available. A breakdown of data requests for the past five years can be seen in Figure 1.

Figure 1. Percent of Inquiries between 2019 – 2023.



As shown in Figure 1, a third of data requests received from 2019 through 2023 fell *outside* the realm of early childhood data, resulting in questions including “how can I register my child in school?” to “who is my legislator?” (This was quite surprising to us. We suspect we get a lot of these inquiries because we land high in Google search results on early childhood topics and, even more so, we have a convenient chat box to ask such questions.) IECAM staff referred those users to other organizations or agencies for answers to questions they could not quickly answer. Thirty-nine percent of inquiries were based on data, ranging from requests about specific slot-gap calculations to percentages of eligible children. Users also requested data about refugees, demographic figures, or maps and visualizations.

Addressing the challenge of ensuring that all users can find answers to their inquiries on our website becomes complex, especially when 30% of the questions are unrelated to data. This presents a formidable task, and there is no one-size-fits-all solution. Catering to the entire spectrum of users is challenging, and, inevitably, we will encounter questions that are not within our scope. It raises another question about how we help state agencies build their resources to provide many of the answers currently being answered by IECAM.

We understand that part of our purpose is to provide support for data education and literacy. Many of our early childhood programs have staff with little working knowledge of data and how to use data appropriately. Some of our state agencies lack capacity in the technology they have to use, as well as the experience of working with their data. A lot of this is out of our control but influences our work with their data, the way we relate to our data sources, and the way we display and describe the data on our website and, more pertinently, on the new Data Hub.

#### **5.4. Managing expectations**

As we started developing this new Data Hub, IECAM took into consideration not only what users said what they wanted from the new Data Hub but also how we can serve a multitude of users with differing preferences, interests, and experiences as well as how we can develop and support greater staff and agency capacity for data and technology. As IECAM moves forward in developing the data hub, how do we manage the differing expectations of various audiences of users? We have always walked a fine line with the differing types of users coming to our website for data and information. The new Data Hub will potentially bring researchers and state entities and agencies to IECAM with more sophisticated methodologies and queries of the data. Yet, it’s also expected that the typical ECEC administrator should be able to use the Data Hub for community needs assessments or to assist in writing grants. How then do we develop a tool meant to be used by such a large swath of individuals with such different skill sets and purposes?

### **6. Conclusion**

We see our path as being one of listening to our users, providing clear and concise descriptions of our data and visualizations for a broad scale of users, and being honest with users about the data they will see on our site, as well as the data they will not see. It is imperative that those of us fortunate enough to work with the data of young children, their families, and programs are honest about not only the benefits that come with accurate and reliable data, but also the limitations that are inherent within the data and the tools that provide and display data.

#### *References*

Pahlka, J. (2023). *Recoding America: Why government is failing in the digital age and how we can do better*. New York: Metropolitan Books.