

BRIEF

MAY 2016

Big Data in Human Services

Introduction

Every day, child- and family-serving agencies collect millions of pieces of data about the children and families they come into contact with and serve. These pieces of data are maintained individually by each agency for internal case management and agency-specific decision-making. Today, child- and family-serving agencies depend on this data to inform decision-making by personnel throughout the organization. However, individual agency data is not complete – each agency has data that only allows for a partial view of the well-being of the families and youth served through the narrow perspective of single agency measures of service receipt and outcomes. The family, youth and children served by state agencies are highly likely to receive services from multiple agencies, both public and private, making it difficult to fully understand the scope of need or the costs associated with addressing these needs through the lens of a single agency.

Due to this narrow lens, it's necessary to move beyond a single agency perspective to serve children and families optimally. For instance, child welfare agencies could benefit from having a fuller understanding of the educational services provided through the schools; the type and amount of any psychotropic medications prescribed; whether there has been any juvenile services involvement; and whether the family is involved with child support, food stamps, or another public service. Likewise, schools should know if a student has child welfare or juvenile services involvement (including changes in placement, which could disrupt the educational environment). This type of information is only available if agencies are willing to move out of their silos and develop relationships with other public and private agencies to share data. To make a transformation of this kind requires willingness by people throughout the public service systems – from the front line workers and educators through the decision-makers at the state level – to embrace the importance of data to effectively administer their duties and to provide and utilize information that is pulled from the data (Webster, Putnam-Hornstein & Needell, 2011).

This concept of data collaboration at the individual level for overall program and systemic improvement is referred to as *Big Data*. *Big Data* is a flashy, and oft misunderstood buzzword, with popular media awash with stories of data breaches and threats to personal privacy – all discussed under the banner of *Big Data*. But what does the term mean in the context of human services? How can we use the information that we collect efficiently, ethically and effectively to serve our populations and improve the overall efficiency of our programs? To answer that question, it is necessary to understand the concept of big data; how our human service systems have developed and have used technology over the years; the laws and regulations that impact the ability to share data across systems; and how to structure data sharing between agencies.

Terry Shaw
Associate Professor

Jill Farrell
Research Assistant Professor
University of Maryland School of
Social Work

Karen Kolivoski
Assistant Professor
Howard University School of
Social Work

This document was prepared for the National Technical Assistance Network for Children's Behavioral Health under contract with the U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Contract #HHSS280201500007C. However, these contents do not necessarily represent the policy of the U.S. Department of Health and Human Services, and you should not assume endorsement by the Federal Government.

Big Data: A Definition

Big Data at the most basic level entails distilling meaning from an incredibly large volume of information generated across a number of varied sources. This is in contrast to the large volume of information contained in only an agency's own data system, which would be considered a single source of data. A single source data system ("little data") does contain a wealth of useful information, but the information contained in these single source data systems is not considered complete. Taking the step to partner with another agency in order to improve the capacity of human service systems to serve client populations means taking the step into *Big Data*.

For the purposes of this paper, a single agency's data system is referred to as "little data" and a collaborative or cooperative system of multiple data systems is referred to as *Big Data*.

The Five V's of Big Data

Veracity: Veracity refers to the accuracy of data and is critical to any data-sharing undertaking. At its core, if the data does not accurately reflect reality, either due to a mistake on the part of the person entering the data or for some other reason, the data loses its practical meaning.

Volume: Volume is the vast amounts of data that are created on any given day. Examples include transaction-based data stored through the years, data streaming in from social media, and sensor and machine-to-machine data being collected through technology.

Velocity: Pieces of data are created and updated at unprecedented speeds and must be dealt with in a timely manner. Reacting quickly enough to deal with data velocity is a challenge for most organizations.

Variety: Data today comes in all types of formats—structured, numeric data in traditional databases; information created from line-of-business applications; and unstructured text documents, email, video, audio, stock ticker data and financial transactions. Managing, merging, and governing different varieties of data is something many organizations will need to grapple with in the future.

Value: Does the analysis of the data and utilization of the information developed from the data produce value for the individuals and agencies involved - and is that value worth the expense? Expense refers not only to the cost of developing and maintaining such a system, but also the increased risk to the individuals whose data forms the basis of this Big Data concept. No matter the size or amount of the data, at its core is information about someone interacting with some system, so the overall value is important to keep in mind.

In 2001, an IT industry analyst named Doug Laney identified the three V's of *Big Data* as "Volume" (the amount of data being collected); "Velocity" (the speed of the data collection process); and "Variety" (the various types of data available; Laney, 2001). Recently, the concept of "Veracity", or accuracy, has been put forth as another component of *Big Data* (Gardner, 2013). The concept of veracity is incredibly important and needs to be considered at every step in the process of utilizing data to inform policy and practice. It has little to do with the size of the data being examined and more to do with how the data is collected, entered and used. Veracity impacts systems in both a *Big Data* and "little data" environment. Finally, a more meaningful component for the human services field is the overall value of the information that can be pulled from the data, which makes "Value" the fifth V of *Big Data*.

Figure 1 shows the relationship between the five V's of *Big Data*. Veracity (or accuracy) is the core from which all of the other components flow. Without accurate data it does not matter how fast, varied or large the data is because its value would be questionable. Each component is expanding, and the value that an agency can gain from the use of *Big Data* also

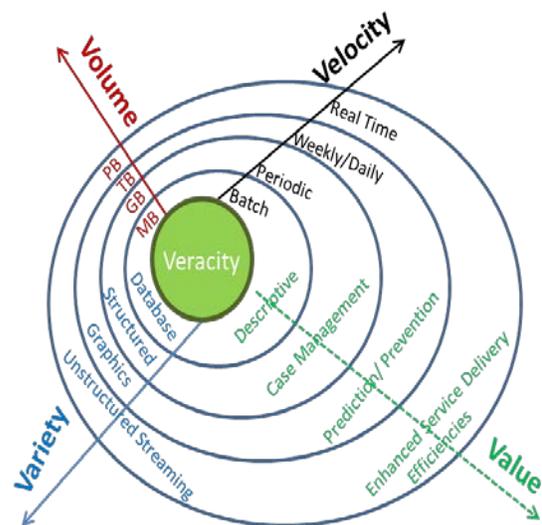


Figure 1: Five V's of Big Data

expands as an agency moves farther along the development of partnerships with other agencies to share information (e.g., moving from an individual agency's "little data" into the presence of more and more varied data – *Big Data*). While there is no doubt that the volume of data and velocity of the data is increasing in human services, the variety of the data remains, for the most part, database-specific. The variety, however, will increase through the utilization of multiple databases.

History of Big Data in Human Services

As computers and information technology became more available in the 1980's, businesses and service providers began exploring ways to utilize this new technology to enhance practice (Born, 1987; Brower & Nurius, 1985; Cnaan, 1989; LeMendola, 1987). As technology has advanced, so has the sophistication of tools available to human services practitioners and educators. In child welfare, the federal government provided funding (up to 75% of the cost of developing and implementing the system) beginning in 1993 for states to develop a State Automated Child Welfare Information System (P.L. 103.66). These are generally compliance-driven case management systems that were not conceptualized with the idea of providing external information to workers. Too often, information systems contain a lot of data, but very little information that is actually useful for working with children and families on the ground. The information that most workers have access to comes solely from within their management system, and often that data has been entered into the system by the worker. Other sources of data that could provide useful practice information (e.g., education system, child support system, health/mental health system, juvenile justice, etc.) are unavailable or challenging to access.

Technology is not the limiting factor at this point; computer systems are more than adequate as far as speed, processing power and general connectivity. Instead, the limiting factors include a lack of understanding of the utility, concerns around resources (time and funding), concerns around data security, legislative concerns, and an inherent hesitation to share data. These are all important areas to address and discuss in each system, but none are insurmountable barriers.

Data Security, Privacy and Legislative Limitations on Data Sharing

A major concern at all levels when discussing the concepts of data sharing and cross-system collaboration has to do with data security and privacy. Controlling information and ensuring that that information is only used as intended are at the roots of data security and privacy (Fitch, 2011). This control of information has both a legal and a trust component, and it is important to address both of these concerns related to data security and privacy to be successful.

Several pieces of federal legislation address issues around data security and confidentiality, including the Health Information Portability and Accountability Act (HIPAA), the Family Educational Rights and Privacy Act (FERPA), the Child Abuse Prevention and Treatment Act (CAPTA), and federal law related to alcohol and drug abuse (42 CFR Part 2). Additionally, jurisdictions often have guidelines at the state, local and/or agency levels that must also be addressed.

Health and Mental Health

Passed in 1996, the Health Insurance Portability and Accountability Act (HIPAA) requires covered entities to ensure the safety and privacy of protected health information (PHI). Covered entities consist of health care providers and their partners who handle billing and records, while PHI includes information that can individually identify someone. Notably, HIPAA allows disclosures "...made pursuant to court or administrative orders..." (HIPAA, 1996).

Education

The Family Educational Rights and Privacy Act (FERPA), originally passed in 1974, is a federal policy designed to protect the privacy of educational records (FERPA, 2014). The release of a student's educational records requires parental consent (or student consent when they are age 18 and older).

Consent for release, however, is not required for the following purposes or to the following entities: (1) school officials with an educational interest; (2) transferring schools; (3) individuals conducting official audits or evaluations; (4) appropriate financial aid organizations; (5) research done on behalf of the school; (6) accrediting organizations; (7) in order to comply with a judicial order; or (8) in case of a health/safety emergency (USDOE, 2014). In 2013, the Uninterrupted Scholars Act (USA) amended FERPA to allow disclosure of educational records to child welfare caseworkers or other representatives without parental consent when child welfare is legally responsible for a child (USA, 2013).

Child Welfare

The Child Abuse Prevention and Treatment Act (CAPTA) originally passed in 1974 and most recently updated in 2010, requires state child welfare agencies that receive federal funding to preserve the confidentiality of all records pertaining to child abuse and neglect maintained by the state (CAPTA, 2014). CAPTA specifies that information related to child abuse and neglect records can be made available to: (1) the subjects of the report; (2) federal, state or local agencies who need the information to protect children from child abuse and neglect; (3) the child abuse citizen review panel; (4) child fatality review panels; (5) judicial systems as necessary for decision making related to the case; and (6) other legitimate entities as determined by the state.

Substance Abuse

The federal Confidentiality of Alcohol and Drug Abuse Patient Records law of 1972 is the law regarding the privacy of an individual seeking help for a substance use disorder and is commonly known as 42 CFR Part 2. 42 CFR Part 2 is designed to protect identifying information of anyone who has received (or is receiving) substance abuse treatment from a federally-assisted program. This law specifies that information related to individuals receiving substance abuse services may be shared if consent has been provided or under several other circumstances including research and evaluation activities if these activities have been approved by an Institutional Review Board for a qualified researcher. Additionally, there needs to be a process assuring that identifying information will not be disclosed.

Juvenile Justice

The Juvenile Justice and Delinquency Prevention Act (JJDP) of 1974 mandates data collection on youth in the juvenile justice system. Data collected includes arrest information, charges filed, court outcomes and probation. It also supports the inclusion of child welfare information and juvenile justice records in formulating treatment plans for delinquent youth. The JJDP specifies that this data can be used in developing data-driven prevention plans, employing evidence-based prevention strategies, and conducting program evaluations to determine impact and effectiveness.

In the context of these federal policies and laws, agencies and systems can still collaborate and share information. Relevant state and local laws likely apply as well. In order to take the first step, agencies can consult their relevant laws and regulations on information sharing and initiate conversations with the appropriate professionals within their agencies who can guide the dialogue about data collaboration and sharing across systems.

Human Service Agency Collaboration

The way that agencies operate their information systems is often described to as silos. Each service system maintains a specific data system to meet the regulatory needs of their system and assist workers in successfully working with clients. Sharing information across systems or silos can be challenging. The Data Quality Campaign (DQC) suggests that there are four barriers to sharing information across systems: turf, trust, technical challenges and time (DQC, 2012). The concepts of turf (who owns the data and has the right to use it) and trust (the willingness to share information with someone else) are at the heart of data sharing and need to be addressed prior to meaningful collaboration. The movement from single system ("little data") to a collaborative or cooperative

system (*Big Data*) can take many forms and require different levels of trust and willingness to share data (turf). Figure 2 shows the four types of collaborative/cooperative structures that can be pursued.

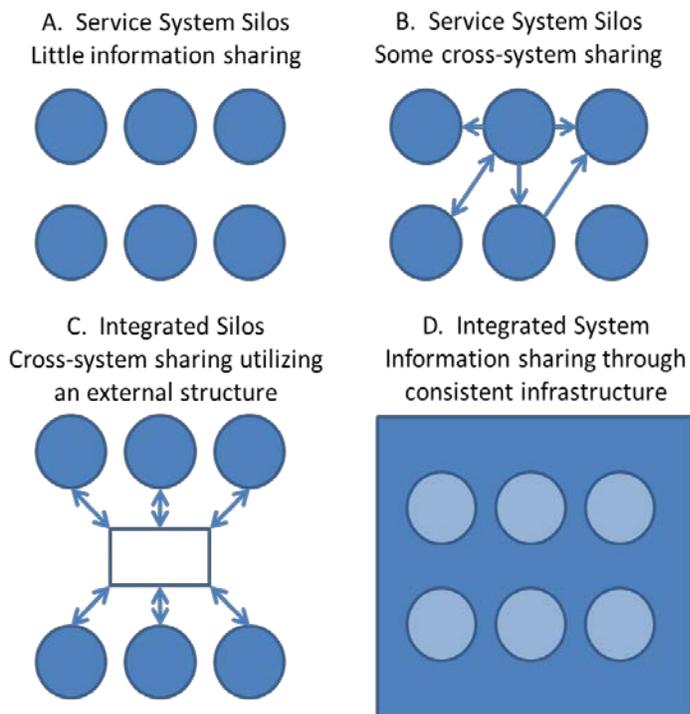


Figure 2: Collaborative/Cooperative Structures

The first type of structure (A) is the prototypical siloed structure that has been the primary configuration of our human services systems since they were initially conceptualized. This is the least collaborative/cooperative structure possible. Each agency's data system sits in isolation from the other systems and there are no mechanisms to share data across agencies. Some information sharing may occur, such as through ad hoc means for program evaluation, through individual worker contact across agencies to verify enrollment, or through the sharing of census-level information. This type of system is common across service delivery systems and the data can be used to effectively manage policy and practice within each system individually (see [Framework for Managing with Data to Achieve Outcomes in Child Welfare: An Overview](#), National Resource Center for Child Welfare Data and Technology for more information).

Each agency maintains a wealth of data ("little data") on the clients that they serve, and this data can and should be used to enhance practice and improve overall efficiency within the agency. But, the information contained in each one of these silos is incomplete as there is no mechanism to readily understand if clients are currently being served (or are eligible to be served) across agencies. And the additional information that exists in other silos might help agencies (if they had the information) to provide more comprehensive and effective services.

The second type of structure (B) remains primarily siloed, but does allow for cross-system data sharing, and possibly integration. Not all human service agencies might be participating cooperatively, but information is flowing between agencies to facilitate decision-making, research and evaluation. This flow might be mainly one way (e.g., education agencies providing information to child welfare agencies, with no information going from child welfare to education) or bi-directional (e.g., cooperative agreements between juvenile services and child welfare) to facilitate program improvements for dually-involved youth. This cross-system sharing might consist of direct agency-to-agency sharing of information at the individual level or could involve agencies contracting with another agency or an outside entity (such as a university) to provide data-linking and analysis for research and evaluation purposes. A cross-system structure like this enhances our ability to understand the fuller context of the lives and needs of children and families being served. For example, cross-systems data sharing can help answer questions related to how many of the children who have exited child welfare come into contact with the juvenile justice system, complete high school, or are from families who are eligible for food stamps, etc. Answering these types of questions allows these agencies to identify populations of need and develop collaborative programs in order to provide necessary services to reduce potentially negative outcomes (e.g., recidivism, involvement in correctional systems, homelessness) and enhance positive ones (e.g., high school graduation, college attendance, workforce participation).

Maryland's Human Services: Service system silos with some data sharing

Maryland is an example of a state that has service system silos with some data sharing. There are a number of data-sharing processes that have been ongoing in Maryland, though not necessarily coordinated across projects. There are various data-sharing structures in Maryland that have been implemented between agencies and between agencies and University partners. Some examples of these structures include:

The Maryland Department of Juvenile Services and the Maryland Department of Human Resources' Social Services Administration share administrative data through the University of Maryland-based Maryland Multi-agency Data Collaborative, (also known as LINKs - Linking Information to eNhance Knowledge), to understand service overlap and facilitate coordination.

The Maryland Department of Education, the Maryland Higher Education Commission, and the Department of Labor Licensing and Regulation share data with the Maryland Longitudinal Data System Center, a state entity housed in the University of Maryland, School of Social Work, to examine educational and work outcomes for Maryland students.

These are just a few of the examples of ongoing data sharing that the State of Maryland has committed to in order to understand cross-systems outcomes and offer further context around the services provided by individual agencies. Other such partnerships and data-sharing agreements exist between state service agencies and partner agencies in Maryland.

Contact: MLDS Center, www.mldsc.gov

LINKs, <http://theinstitute.umaryland.edu/frames/links.cfm>

The third type of structure (C) consists of integrated silos in which cross-system sharing occurs through the utilization of an external structure (such as a third party platform providing data linking and information sharing opportunities). Agencies maintain their siloed arrangement and utilize an external structure to operate and maintain a cross-systems data linking and access infrastructure. The external structure can be operated by a state agency, a third party vendor, or a university partner. The integrated silos approach allows for the bridging of disparate agency administrative data in a timely manner that can be used in developing healthcare passports, truancy reports, medication management systems and other possibilities designed to get meaningful and actionable information to human services staff as soon as possible in order to intervene for children and families. This structure requires clear identification of the data that can be shared by each agency and with whom this information can be shared. Data from each agency is uploaded daily (or as often as the agency agreement allows), linked using an agreed upon data-matching algorithm, and made available to all participating agencies based on the agreements in place. Access to the data for participating agencies is possible through a secure web-based

interface. This structure is primarily a means of enhancing case management and service provision through increased understanding of the needs of client populations. An evaluation or research plan needs to be in place and agreed upon by participating agencies in order to create the necessary files to understand trends, evaluate programs and make policy recommendations.

Finally, the fourth type of structure (D) is shown as an integrated system in which information sharing occurs through a consistent infrastructure and network of agreements allowing for information to be accessible to all agencies in the integrated system. An integrated system has a uniform means of identifying individuals regardless of the service system they interact with and, equally important, there is a way for human services workers to access all pertinent information about a client to enhance the ability to provide meaningful services. For example, this level of integration allows a caseworker who is helping to enroll a client into food stamps know if the client is eligible for Medicaid and can facilitate the process to enroll the client into that service. Additionally, if a child welfare worker is investigating a family for neglect and finds that there is a need for

Miami-Dade County, Our Kids: Using an external structure to facilitate collaboration

Our Kids of Miami-Dade/Monroe, Inc. has developed a data-sharing system along the lines of an electronic passport called "Follow My Child." This electronic passport utilizes a data warehouse maintained by an external group (Our Kids of Miami-Dade/Monroe) with relevant information from various agencies to provide required information to case workers and foster parents on children they serve.

The development of this system required extensive review of federal and state data sharing and privacy laws. These reviews led to discussions with each agency in order to develop the requisite Memoranda of Understandings (MOUs). Currently, Our Kids has MOUs with the Florida Department of Children and Families, Florida Department of Juvenile Justice, statewide juvenile judges, and Miami-Dade County Public Schools. These MOUs allow data sharing in order to change practices and make children and families safer.

Contact: Our Kids, www.ourkids.us

health insurance and therapeutic services, these services can be applied for and scheduled immediately through the integrated interface.

Although this fourth type of structure — a fully integrated and operational data-sharing system — is ideal from a data-sharing perspective, the reality is that this type of system is currently beyond reach both legally and fiscally in most locales. Additionally, many agencies are relatively new to realizing the potential of sharing data across systems to benefit clients and agencies. Through working toward mutual goals and manageable data-sharing partnerships, these relationships can gradually build toward increased and improved integration and data collaborations, moving from the siloed approach (A) into some form of data sharing (B) or (C) to enhance practice and service delivery.

Conclusion

Big Data is here. We live in an era where more and more information is maintained in electronic information systems and we are at the critical point now where ethical data sharing can provide empirical evidence to inform day-to-day practice as well as expanding knowledge and research. Given the many privacy and legislative concerns, careful and thoughtful partnerships across agencies can and should occur to better serve clients in human services systems. Many options exist for creating partnerships across systems for data sharing and ongoing collaboration. Although effort needs to be made toward adhering to regulations and respecting agency partnerships, the resulting efforts of serving clients in the present day, and well into the future, provides reward for such endeavors. The era of *Big Data* is upon us, and, it is worthwhile to embrace the possibilities.

Montgomery County, Maryland Integrated system structure

Montgomery County, Maryland has been steadily moving toward an interoperable system across child- and family-serving agencies since 2008. The county structure is conducive to this type of system as there is a single director overseeing all of the health programs. This allows for centralized administrative functions within the health programs, which Montgomery County believes is the hallmark of an integrated and interoperable system.

The move toward this integrated structure began with the desire to have an Electronic Health Record (EHR) and seamless customer service across all public agencies (“no wrong door” access to services for families and children). The process to develop the confidentiality and data sharing agreements with the county attorney took over 18 months, but now the confidentiality framework is in place allowing role-based access to cross-system information.

The Electronic Content Management System contains real time interfaces with most systems and non-real time interfaces with child welfare. All agencies have developed an agreed upon common face sheet in which service needs are identified and referrals made electronically and immediately through the integrated system.

Contact: Montgomery County, Maryland,
www.montgomerycountymd.gov/hhs/

References

- Born CE. (1987). Microcomputers and field instruction. *Journal of Social Work Education*, 23(2), 69-74.
- Brower AM & Nurius PS. (1985). A teaching model for the use of computers in direct practice. *Computers in Human Services*, 1(1), 125-131.
- Child Abuse Prevention & Treatment Act (CAPTA)
<https://www.childwelfare.gov/pubs/factsheets/about/>
- Cnaan RA. (1989). The Impact of Information Technology on Social Work Practice.
- Data Quality Campaign. (2012). Pivotal role of policymakers as leaders of P-20/Workforce data governance. <http://dataqualitycampaign.org/resource/pivotal-role-policymakers-leaders-p-20workforce-data-governance/>. Retrieved 9 June 2014.
- Drug abuse (42 CFR Part 2) <http://www.ecfr.gov/cgi-bin/text-idx?rgn=div5;node=42%3A1.0.1.1.2>
<http://www.samhsa.gov/about-us/who-we-are/laws/confidentiality-regulations-faqs>
- Family Educational Rights and Privacy Act (FERPA)
<http://www2.ed.gov/policy/gen/guid/fpco/ferpa/index.html>
- Fitch D. (2011). Controlling their story: Protecting the privacy of foster care youth. *Child Welfare 360: Child Welfare and Technology*, University of Minnesota.
- Gardner D. (February 22, 2013). The Open Group Panel explores how the big data era now challenges the IT status quo. The Open Group. <https://blog.opengroup.org/2013/02/22/the-open-group-panel-explores-how-the-big-data-era-now-challenges-the-it-status-quo/>
- Health Information Portability and Accountability Act (HIPAA) <http://www.hhs.gov/hipaa/>
- The Juvenile Justice and Delinquency Prevention Act (JJDP) of 1974
<http://www.juvjustice.org/federal-policy/juvenile-justice-and-delinquency-prevention-act>
- Laney D. 3D data management: Controlling data volume, velocity and variety. Gartner.
<http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>. 6 February 2001.
- LeMendola W. (1987). Teaching information technology to social workers. *Journal of Teaching in Social Work*, 1(1), 53-69.
- National Resource Center for Child Welfare Data and Technology (NRC-CWDT)
<http://www.acf.hhs.gov/programs/cb/resource/nrc-cwdt>
<http://www.nrcdr.org/assets/files/NRCDR-org/Framework-Data-Overview.pdf>
- Uninterrupted Scholars Act (USA) <https://www.congress.gov/112/plaws/publ278/PLAW-112publ278.pdf>
- Webster D, Putnam-Hornstein E & Needell B. (2011). Using data for child welfare system improvement: Lessons learned from the California Performance Indicators Project. *Child Welfare 360: Child Welfare and Technology*, University of Minnesota. Spring 2011, pp. 6-133.

ABOUT THE NATIONAL TECHNICAL ASSISTANCE NETWORK FOR CHILDREN'S BEHAVIORAL HEALTH

The National Technical Assistance Network for Children's Behavioral Health (TA Network) operates the National Training and Technical Assistance Center for Child, Youth, and Family Mental Health (NTTAC), funded by the Substance Abuse and Mental Health Services Administration, Child, Adolescent and Family Branch. The TA Network partners with states, tribes, territories, and communities to develop the most effective and sustainable systems of care possible with and for the benefit of children and youth with behavioral health needs and their families. The TA Network provides technical assistance and support across the country to state and local agencies, including youth and family leadership organizations.