ACTIONABLE INTELLIGENCE FOR SOCIAL POLICY

METROLAB NETWORK

NNIP NATIONAL NEIGHBORHOOD INDICATORS PARTNERSHIP

TECHNOLOGY FOR CIVIC DATA INTEGRATION

SUPPORTED BY

THE ANNIE E. CASEY FOUNDATION

Table of Contents

A digital version of this report with resource web links can be viewed and downloaded as a PDF here: **http://bit.ly/2xaQW2L**
Throughout the document, **orange text** indicates an internal link to the Appendix section.

Introduction

Over the past decade, there have been concerted efforts by public agencies, universities, and nonprofits to put the troves of administrative data that governments collect to greater use in making better decisions and improving the quality of life for families and communities. As the Annie E. Casey Foundation defines it, "integrated data systems (IDS) [are] an information tool which a growing number of leaders are leveraging to enhance coordination across a wide range of social programs that have a shared stake in supporting families. By connecting regular public agency records, integrated data systems enable policymakers, practitioners, advocates and researchers to look across programs and add critical depth and details to complex social issues." [1]

Administrative data sharing and linkage can be a powerful and cost-effective tool to inform policy and practice. It is essential that data sharing and integration be done within a carefully constructed governance process and legal framework, with appropriate technology to safeguard data and staff capacity to manage and link data across multiple agencies.

Local jurisdictions are developing IDS to facilitate the integration and use of personally identifiable information for safe and ethical use.[2] For over a decade, Actionable Intelligence for Social Policy (AISP) and National Neighborhood Indicators Partnership (NNIP) have supported and convened IDS to seed new efforts, identify and develop best practices, and increase access to the insights IDS contain. Counties, cities, and states that have successfully built IDS have shown that the biggest challenges sites encounter are political, relational, and procedural rather than technical. Discussions of ethical use, governance, legal issues, privacy, and data standards are critical, and we encourage readers to use existing resources from AISP, NNIP, and MetroLab.

RESOURCE

You can learn more by visiting the Annie E. Casey Foundation's blog article ***Improving Child and Family Services Through Integrated Data Systems.***

RESOURCE

For a more in-depth explanation, please read AISP's ***The Integrated Data System Approach: A Vehicle to More Effective and Efficient Data-Driven Solutions in Government*** and NNIP's ***Using Integrated Data to Improve Communities: Lessons from a Cross-site Project***.

Nevertheless, once these adaptive challenges are addressed, questions may remain regarding which technical approach to integration is optimal and how to build a system that can be adapted to incorporate advances in technology over time.

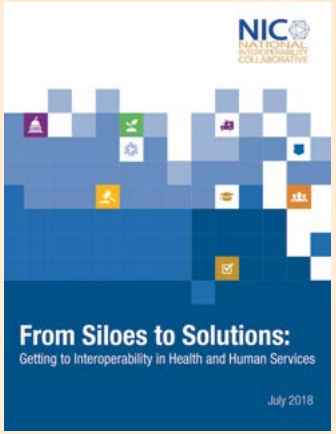
The purpose of this report is to describe key considerations in building and sustaining IDS and the various technology approaches that may be helpful in overcoming challenges in data integration. Sharing, integrating,[3] and using administrative data to inform policy and practice is technically complex, and involves navigating myriad constraints around cost, architecture, storage, access, and staffing—just to name a few.

Heterogeneous approaches to data infrastructure, transfer, linking, and security are to be expected in an emerging field, but both established and developing efforts would benefit from cohesive guidance regarding the technical considerations of data integration, as well as a common understanding of the range of options that can be weighed based on context-specific restrictions. The aim of this report is to provide guidance to better equip states, local governments, and universities with technical approaches for data sharing and integration.

This report is primarily intended for government agencies and their partners who are working on using integrated data to improve services, drawing from NNIP and AISP's long-standing leadership in those fields, and developed as part of MetroLab's Data Science and Human Services Lab.

[3] INTEROPERABILITY VS INTEGRATION:

"Interoperable" systems exchange information in real time, often to inform decisions about individual client care. Electronic health record systems and "no wrong door" eligibility screening for social programs are examples that employ interoperable systems. (For more, visit the **Stewards of Change Institute**.) "Integrated" data systems match individual records from different data sources and are not often used for real-time care coordination, but rather for analyzing information across agencies to inform better policy decisions, evaluating the benefit and cost of programs, and reallocating resources (For more: **Policy Labs**, **AISP**). These different use cases present different challenges and lead to different technical architectures. They are not mutually exclusive; in some cases, "integrated" data systems may be "interoperable" with other systems or software.



For additional information, we recommend reviewing NIC's ***From Siloes to Solutions: Getting to Interoperability in Health and Human Services.***

Audience and Purpose

This report is intended to provide brief technical guidance for jurisdictions that are looking to develop, implement, and/or upgrade their IDS efforts. It could also serve as a tool that agency leadership and policy makers utilize to foster discussion on data integration capabilities within and across departments. It is structured around eight "key considerations." For each consideration, example technologies and approaches are discussed along with a series of questions that organizations building or operating IDS's should ask themselves.

The factors determining how and when to implement an IDS will be unique to each organization. This report is meant to serve as an initial starting point, and the authors of this paper highly recommend seeking out additional resources and field experts as your jurisdiction commences and/or continues to have conversations around this effort. In particular, it is imperative to have a solid understanding of your data integration needs prior to designing and procuring an integrated data infrastructure. This will increase the likelihood of adopting an infrastructure that has a positive impact on your organization's practices and outputs. Resources such as the **Center for Data Science and Public Policy's Data Maturity Framework** or **Civic Analytic Network's Data Maturity Self-Assessment Tool**, are available to help guide you as you examine your data.

Finally, because this is a fast-evolving field, we recommend readers refer to the latest guidance available from other organizations working on these issues.

We would like thank the following group of data scientists, policy advocates, and public leaders for being so generous with their time and expertise:

Joy Bonaguro, City and County of San Francisco

Matt Gee, Brighthive

Lisa Green, Data Domino Lab

David Hill, UNC Charlotte Urban Institute

Bill Howe, University of Washington

Anjum Khurshid, University of Texas at Austin

Julia Koschinsky, Center for Spatial Data Science

Jason Lally, City and County of San Francisco

Graham MacDonald, Urban Institute

Christopher Mader, University of Miami

Kathy Pettit, National Neighborhood Indicators Partnership

Deepthi Puram, University of Miami

Matthew Tamayo-Rios, Open Lattice

Emily Wiegand, Chapin Hall

Bill Yock, Santa Clara County

An additional thanks to the Advisory and Coordinating Group who helped in the development and execution of this paper.

Natalie Evans Harris
Harris Data Consulting (and lead author)

Stefania Di Mauro-Nava
MetroLab Network

Amy Hawn Nelson
Actionable Intelligence for Social Policy

Leah Hendey
National Neighborhood Indicators Partnership

Christopher Kingsley
Annie E. Casey Foundation

MetroLab Network, Actionable Intelligence for Social Policy (AISP), and the National Neighborhood Indicators Partnership (NNIP) would like to thank the Annie E. Casey Foundation for their generous support in the development of this report. We acknowledge that the findings and conclusions that are presented in this report are those of the authors alone, and do not necessarily reflect the opinions of our contributors, nor the Annie E. Casey Foundation.

Building an IDS: Key Considerations

An integrated data system (IDS) collects, links, organizes, and delivers data about clients' needs, risks, interventions, and outcomes. Building and operating an IDS is a complex data integration process that takes the raw source data and manipulates it to draw meaning from the integrated data. With so many technology choices, there are many dimensions to consider, and asking the right questions on what and how to procure, develop, and operate is essential to the use and sustainability of the IDS. There are five key phases to the data integration lifecycle. Data integration is an iterative process and many organizations enter and exit the phases repeatedly:

- **collecting the data from two or more sources,**
- **validating the data to ensure data are accurate,**
- **normalizing and standardizing the data to maintain integrity,**
- **linking and appropriately anonymizing the data,[4] then**
- **analyzing and disseminating the data.**

Successful adoption of IDS is critical to state and local governments' ability to deliver successful and cost effective social services.[5] While this paper will focus on the technology implications, there are plenty of legal, policy, and financial[6] considerations that should also be examined.

[4] HOW IS APPROPRIATE ANONYMIZATION DETERMINED?
De-identification and anonymization standards should be set by participating data partners.

5
RESOURCE

Pew Charitable Trusts' report, ***How States Use Data to Inform Decisions***, outlines some of these reasons and is a helpful guide on this topic from a policy perspective.

6
RESOURCE

There are a number of resources you can view to learn more about these topics. We recommend, **State Data Sharing** for legal suggestions, **NACo** for some policy ideas and **Stride NYC** to see a financial example.

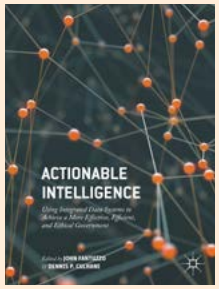
There are different options for IDS system architecture that Prashant Kumar covers extensively in "An Overview of Architectures and Techniques for Integrated Data Systems Implementation," within *Actionable Intelligence: Using Integrated Data Systems to Achieve a More Effective, Efficient, and Ethical Government* (2015).[7]

A key consideration when adopting software is its licensing model: broadly open source versus proprietary licensed software. Open source software is free to procure, but working group members point out that open source tools could require a higher technical capacity within staff to support and maintain the software. The sustainability of the tool should also be consideredgiven that, unlike with proprietary software, no third party will be required to maintain and enhance open source tools over time. To mitigate this challenge, look for well-supported open source tools—those where there may be a commercial offering built on top of open source licensed software or where there is a robust community supporting its development.

Which technologies are available to meet your data integration infrastructure needs and what is right for your organization? There are many technology options available; below are examples of options and each of the following sections lists what the technologies can and cannot do. An overarching consideration that is often overlooked when considering technologies are the labor costs, procurement costs, and maintenance costs associated with running an integrated infrastructure. As such, this paper includes questions under each key consideration to illuminate costs and benefits.

[7] ORIENTING IDS DESIGN FOR DISTINCT USAGE SCENARIOS

Architectural approach for leadership- and research-oriented use cases	Architectural approach for practice-oriented use cases
• Designed to help uses gain "insight" into problems at the population level • Designed to support deep analytics • Completeness of data has greater emphasis • Client identity is generally kept confidential • Document and image sharing is generally not needed • Partitioning of data for compliance is not required	• Designed to help users gain "situational awareness" of the needs and risks at the client level • Designed to support broad information sharing and presentation • Timeliness of speed of data has greater emphasis • Client identity is made available to authorized users • Document or image sharing may be required for care coordination • Partitioning of certain types of data may be required for regulatory compliance



Full details and additional information can be found in AISP's book, ***Actionable Intelligence: Using Integrated Data Systems to Achieve a More Effective, Efficient, and Ethical Government.***

Consideration 0: Staffing Expertise

Regardless of which technology options are chosen, organizations require staff with substantive business expertise regarding the issues, programs, and context-specific social service areas. No technology can replace this necessary contextual knowledge. In addition, maximizing return on investment in an IDS requires staff with domain-specific data analysis experience in the subject matter. During a 2017 discussion with AISP and NNIP network members, many stressed the importance of hiring for interdisciplinary skills. It is important to have individuals that know at least one programming language and have an ability to learn others. It is critical that they maintain that flexibility to employ the right tools or approaches nimbly, rather than fitting a challenge narrowly to their own skillset. Many program administration staff, (ie. social services, behavioral services, public health, education, public safety and justice, environmental protection, etc.) are being cross trained to learn data analysis and visualization skills. There is a wide variety of training options available to users to learn the critical data science skills necessary. Many universities8 offer Data Science degree and certificate programs including interdisciplinary curriculums to meet the growing need. We have also included an appendix of job descriptions with this paper as a resource.

Each working group member stressed that people are critical to successfully building and operating an IDS. Regardless of which data integration and data analysis infrastructure technologies are chosen, an organization must have a highly capable staff to maximize the investment. The infrastructure and technology choices put in place must integrate tightly with the underlying work culture and business processes. Based upon conversations with AISP and NNIP members in 2017 and interviews for this paper, four key skill areas emerged, each requiring a different mix of roles and responsibilities, including:

- **Data Storage and System Administration:** solution architects, system administrators, security administrators, database administrators, production support operators, etc.
- **Data Integration:** data / information architects, data modelers, data engineers, integration/ interface engineers, etc.
- **Data Analytics**: research / evaluation specialists, data analysts, data scientists, business intelligence analysts/programmers, etc.
- **Data Publication:** website administrators, UX designers, BI portal administrators.

When thinking about staffing expertise, ask these questions:

1. Are the technology solutions chosen on-premise, cloud-based, or a combination? Are existing staff able to support both on-premise or cloud-based systems, or are new staff needed?

2. Are the data highly integrated and aggregated or just a collection of unstructured data sets? Will Data Analysts need to help in modeling, prepping and loading data or can they be self-sufficient using the technology infrastructure chosen?

3. Do you have a pipeline/efficient process for recruiting critical skills? Does your local employment market have a shortage of skilled resources?

4. Will new staff be needed to maintain the new technology infrastructure? Can it be contracted out?

5. Will new or existing staff need training to use the new technology infrastructure? Does the vendor offer training?

8

RESOURCE

There are a number of sources online to learn more about these programs. A good starting point is **Data Science Community's List of Degrees by Program**.

Consideration 1: Data Management

Data management is the process by which leaders in the organization establish strong data governance practices[9] that protect the quality of data throughout the entire data integration process. It is also essential in adding context to the integrated data to facilitate appropriate research and analysis. Establishing the business processes necessary to make sure the key factors discussed in this paper are realized is integral to creating a culture of accountability for managing risk. The Administrative Data Research Facilities Network (ADRF) drafted an in-depth paper on data quality and standards[10] that dives into the many complex aspects of managing data effectively. Metadata is imperative to keeping up with the history of source data after it has been integrated. There are many different types of metadata[11] and organizations must define what they want to collect upfront. There are initiatives and technologies available to automate many aspects of metadata management[12] including **Informatica**, **IBM Infosphere**, and **Azure Data Catalog**. Additionally, data modeling tools like **ERWin** or **ERStudio** are used to manage metadata definitions.

RESOURCE 9

We recommend reviewing NNIP's **Resource Guide to Data Governance and Security**.

RESOURCE 10

Additional information can be found in "**ADRF Data Quality and Standards**," (O'Hara et al., 2018).

[11] SO, WHAT ARE METADATA?

In short, metadata are data that provide information about other data. Basic information always collected includes: purpose of the data, time and date of creation, creator or author of the data, location where the data was created, standards used, file size, source of the data, and process used to create the data.

[12] REAL WORLD EXAMPLE
SAN FRANCISCO'S DATA GOVERNANCE API:

The City and County of San Francisco created DataSF, which is responsible for overseeing the city's data practices. **DataSF** has created a number of reports, publications and guidelines to help the community better navigate data sources within its jurisdiction and provide transparency to the community. As part of this effort, DataSF maintains a public facing API guidebook, the **Open Data Developer Resources**, in an aim to have developers and others who would like to interact with the city's data meet their standards. In addition to API data governance, DataSF has a number of other resources available, including Metadata Standards, Strategic Planning documents and data quality resources, which can be accessed on their **website**.

We recommend addressing issues of data standardization as far upstream as possible. When interacting with data partners, attempt to clearly articulate and document the required data elements and formats, if possible aligning them with industry standards and regulatory guidance. A standardized data collection process will result in a data infrastructure that is less complex and easier to maintain.[13] From a practical perspective, this means engaging the IT stakeholders and all data partners as early in the process as possible, and building a relationship of trust and communication. With that said, this is often an intense effort depending upon the complexity of the data environment, as evidenced by the healthcare industry. Despite very well-developed data standards like **LOINC**, **SNOMED**, **ICDs**, the integration of data from different healthcare sources is still a huge challenge. Moving downstream, if possible, you may try to affect greater alignment on reporting and accountability from government agencies and funders. This would enable shared procurement to reduce costs and increase interoperability, result in improved governance and streamlined legal and political processes, and reduce redundancy.

There are initiatives and technologies available to automate many aspects of data management including **Informatica** and **Socrata**. The ideal technologies leverage tools such as **application programming interfaces (API)** to collect information about the transformation of data automatically.

When thinking about data management, ask these questions:

1. What metadata are collected about data provenance, quality, processing history, etc?

2.How are metadata collected? Where are they stored? How are they accessed?

3. Does the solution support monitoring of data collection, cleaning, and integration activities? How does data quality feedback get passed to the data generators, and how does one track reporting deadlines and data completeness?

4. How does the infrastructure support good governance practices?

[13] REAL WORLD EXAMPLE: PCORNET COMMON DATA MODEL (CDM)

In 2014, the Patient Centered Outcomes Research Institute (PCORI) invested more than $250 million in the development of PCORnet. PCORnet is a large, highly representative, national "network of networks" that collects data routinely gathered in a variety of healthcare settings, including hospitals, doctors' offices, and community clinics. By engaging a variety of stakeholders—patients, families, providers, and researchers—PCORnet empowers individuals and organizations to use data to answer practical questions that help patients, clinicians, and other stakeholders make informed healthcare decisions. The Network uses a Common Data Model that does not include any real patient/member/participant or provider identifiers. Instead, each site within a network contributing data creates a pseudo-identifier for each patient and provider. This pseudo-identifier is not traceable to a specific individual without a separate crosswalk table that each site stores separately from the data and is not part of the Common Data Model. You can find more information on PCORnet's **website**.

Consideration 2: Security & Privacy

One of the biggest challenges with integrating administrative data is ensuring that the integrity of the data and individual privacy are legally, technically, procedurally, and physically maintained throughout the process. From a technology perspective, it is imperative that data be protected in accordance to agreed-upon standards, regulations and laws. For research and analytic efforts, this often means leveraging anonymization techniques such as **data masking, data aggregation**, and **data obfuscation**. It also means ensuring proper consent to use administrative data. Many reports have been written regarding ethical principles and guidelines for research, including the Belmont Report (1979), a seminal report on this topic.[14] AISP's series of expert panel reports provides extensive guidance on how to responsibly and securely address privacy concerns associated with utilizing administrative data for evaluation and analysis.[15] For data integration, it is important to ensure that a strong encryption algorithm is deployed (e.g. **AES**, **DES**, **RSA**, & **ECC**) and that the system administrator can manage the encryption keys which keep the data private. It is important to note that data should be encrypted while in transit and at rest and can be applied at the storage layer and sometimes at the database layer of an infrastructure.

When thinking about security and privacy, ask these questions:

1. Does the solution comply with Federal, State and Local data protection requirements, such as **FedRAMP compliance**, **FERPA**, and **HIPAA**? This may impose constraints on the underlying infrastructure (such as using government clouds rather than public clouds), on encryption, and many other aspects of the system.

2. How does the solution manage user accounts to ensure authorized access to the correct data, at the right times, for the right reasons? Is there a single sign-on solution? Are there **role-based access controls (RBAC)**?

3. How is access logged, and what are the audit trails?

4. How are data encrypted at rest and in transit? What is the strength of encryption used? How are encryption keys managed?

5. Does the solution de-identify personal information? Are the data stored with anonymized identifiers, or are they aggregated? Are the anonymized identifiers sufficient to combat the "mosaic effect?"[16]

6. How do staff verify results of data privacy and security? Are logs reviewed frequently?

RESOURCE 14

The Department of Health and Human Services' Belmont Report (1979) discusses the basic ethical principles and guidelines for human subjects research. See Matthew Salganik's ***Bit by Bit: Social Research in the Digital Age*** for a nuanced discussion of this topic.

RESOURCE 15

We encourage you to review all four of **AISP Innovation Expert Panel Reports**.

[16] **MOSAIC EFFECT** is the ability of individuals to derive the identity of individuals from distinct/anonymized pieces of information. Mosaic effect is an important concern in security and privacy that practitioners and agencies should know about.

Consideration 3: Data Collection

In order to make decisions about which technical architecture is most appropriate, governments should first seek to understand the characteristics of the data and how uses will interact with them. In particular, characteristics such as data format, size, quality, and purposes of collection have tremendous implications for how the infrastructure accepts data from the source systems. Based upon your assessment of the source data, it is important to determine how the source data will be collected and uploaded into an integration platform. While there are many ways to upload data, this paper categorizes them as either manual (an individual uploads the data) or automatic (there is a connection between the data source and the integration system). In many infrastructures, a hybrid approach that allows for some manual uploads and some automated connections is implemented.

- **Manual Upload:** There is no direct connection between the integration infrastructure and the source system. The source data are copied from the data owners system and manually uploaded by the receiver to an integration infrastructure. Staff creates scripts to organize and prepare the data for linking. For example: **The Institute of Social Capital** at the University of North Carolina, Charlotte, works with community partners to upload their data onto USB or other collection hardware, and then review, format and synthesize that data based on a very specific set of criteria imposed by their data sharing agreements and information requested by researchers.

- **Automated Collection:** The infrastructure conducts all collection activities in accordance with a pre-defined data model. Infrastructure is often connected to the source system, allowing for near real-time collection and validation of data. Staff focuses on validating the results. For example: Allegheny County, in western Pennsylvania, has one of the most robust integrated **data collection systems** in the country. Its Department of Human Services integrates data from both internal and external sources to create a data warehouse that helps inform most of its efforts, including a **screening tool for child welfare**.

Today, an IDS must be dynamic and agile enough to support new requirements, changing schemas, and new data streams. Unfortunately, the lack of broadly-accepted standards for administrative data collection means that IDS must often accommodate unstructured or heterogeneous data from multiple data sources and standardize that data for analysis. How the infrastructure addresses data normalization and linkage can vary from warehousing the raw **unstructured data** so that staff can run ad hoc scripts to clean and normalize the data as necessary, to fully integrated data models and **Extraction/Transformation/Loading (ETL)** routines for the system to automatically structure and ingest the data into an analytics schema. The software tools leveraged in the data infrastructure must reflect, and regularly define, the relationship between the raw data and the people who use it.

There are many legal, technical, and financial questions that drive how source data are brought into the integration infrastructure. Understanding these factors will drive how data from separate sources are brought into your integration infrastructure and will determine the role of technology and staff.

When thinking about data collection, ask these questions:

1. What is the cost of establishing the connection and implementing technology necessary to automatically collect and clean the data?

2. What are the legal barriers to establishing a connection?

3. How do your data partners share their data? Via spreadsheets and shared drives, via API, or via some other mechanism?

4. What business and technical factors go into creating a standardized data model? (The data model implements the business and technical requirements necessary for the data to be processed.)

Consideration 4: Data Storage

The decision to store data on-site, in a data center, or leverage the many cloud-based solutions available (**Microsoft Azure**, **Amazon Web Services**, etc.) is largely dependent upon cost and data protection requirements for the source data. The IDS will most likely require a warehousing solution to properly store and make data accessible. Data warehousing has been the most common answer to data analytics development for decades and has thus become a large and entrenched industry.

- **Repository.** The simplest warehousing options store the source data as **dimensional data** in a repository. The main function of the repository is abstracting across multiple data sources. All integration processes are conducted by staff who create scripts to extract, transform, and load aggregated, longitudinal integrated datasets. Data warehouses typically require some type of database management system (Oracle, SQL Server, etc.).
- **Data lake.** Due to the increasing complexity of data sources, many organizations have moved to data lakes, which are a system of repositories. The data lake can catalog and create curated dataset(s) for staff to review and contextualize. Data lake technologies includes **Informatica Enterprise data catalog** and **Azure Data Catalog**.
- **Non-relational database.** A more ambitious and still evolving strategy relies on **non-relational (or NoSQL) databases** with indexing and tagging at the front-end. This allows greater flexibility and faster querying capabilities for large databases.

The increased acceptance of cloud-based solutions has meant a move from monolithic warehouses and administrative databases toward more modular data services and warehousing strategies. **Data vaults** are a new warehouse methodology that organizes data into hubs, links and satellites to make it easier to incorporate new data sources. Many organizations leverage existing database technology choices as part of enterprise licenses or subscriptions. Technology options include **Wherescape**, **Snowflake** and **Denodo** to create and manage warehouses and lakes.

When thinking about data storage, ask these questions:

1. If using a cloud storage provider, is it up-to-date on data center and industry certifications such as **HIPAA**, **FedRamp**, **PCI DSS**, **SSAE 16**? What happens if there is a data breach?

2. What is your backup strategy? How will the data be backed up? Is there redundancy?

3. Are there bandwidth limitations for uploading or retrieving data? Restoring data? Transferring data?

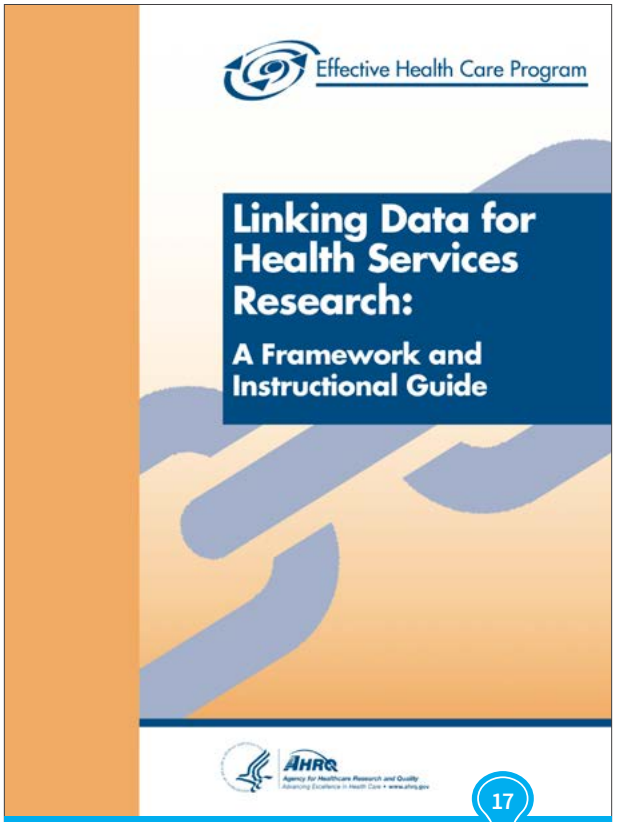
4. What common data analytic infrastructures being used?

5. Can you leverage existing data warehouse technologies?

6. How are the common data analytic infrastructures licensed?

Consideration 5: Data Linking

Data linking is the process of integrating different data sources, based upon common business keys and other identifying information (i.e. SSNs, names, addresses, medical record numbers, etc.). There are two key methods for data linkage: deterministic matching and probabilistic matching (also known as fuzzy matching). Deterministic matching is more precise since it looks for exact matches in the content and format of datasets (i.e., identical SSN) whereas probabilistic matching looks for closeness in the data (i.e., identical SSN with or without dashes) and provides weighted scores for likelihood of matching. Given the requirements of exact matching, when dealing with large datasets, deterministic matching can be easier to implement but also more resource intensive if data must be formatted and cleaned in advance of matching. Fuzzy matching is more commonly used, since fuzzy matching provides weighted scores. The algorithms implemented can be complex depending upon the data sources and the risks factors associated with fuzzy matching are pre-determined by the user. There are a number of technology solutions[17] for data linking. **Choicemaker** is an open source technology that leverages fuzzy linking; **SAS** and **Informatica** are common commercial tools for linking data in a variety of ways.



RESOURCE 17

For additional information of data linking software options, we recommend the Department of Health and Human Services' **"Linking Data for Health Services Research: A Framework and Instructional Guide"**

When thinking about data linking, ask these questions:

1. Does the technology support real-time data uploads, routine data uploads, and ad hoc data uploads? What is the acceptable cadence for batch processing (weekly, monthly, quarterly, yearly, etc.)?

2. Is there a limit to the number of records the software can link? Does your data linkage software license cap the number of integrations? Does performance degrade based upon a large number of records?

3. How is deduplication and entity resolution handled? Can the software or approach do fuzzy linkage[18] between individuals or organizations based on multiple fields, or is there a single uniform identifier scheme? Does the linking software present false-positives and false-negatives for inspection and resolution for human review? Is there automated workflow and/or notification to data stewards to review and reconcile?

4. What happens when a new data source is introduced and the matching rules need to change?

5. How will quality control and cleaning be handled? Do you rely on your data partners to clean and normalize data, do you attempt to automate the process, or do you expect your analysts to clean it manually?

6. How are infrastructure changes made? If a vendor provides support, do they need to be involved in every change, and if so, what are the expectations around timeline and cost?

[18] **PROBABILISTIC/FUZZY VERSUS DETERMINISTIC/ABSOLUTE MATCHING:** Deterministic or Absolute Matching mainly looks for an exact match between two pieces of data. It is ideal if your source systems are consistently collecting unique identifiers like Social Security Numbers to make the matches. On the other hand, Probabilistic or Fuzzy Matching uses a statistical approach in measuring the probability that two records represent the same individual. It is designed to work using a wider set of data elements to be used for matching. It uses weights to calculate the match scores, and it uses thresholds to determine a match, non-match, or possible match. Fuzzy Matching also leads to Type I and Type II errors which need to be taken into consideration. (A Type I error is the rejection of a true null hypothesis (also known as a “false positive” finding), while a Type II error is failing to reject a false null hypothesis (also known as a “false negative” finding).)

Consideration 6: Data Access and Dissemination

When considering the technologies will best support data integration, it is important to consider 1) planning for how the integrated data will be accessed by analysts and 2) how the data can be disseminated in a way that maximizes the value of the integrated data. These considerations are too often over-looked. Secure data access for analysts can be approached in a variety of ways, including but not limited to VPN remote access, limited and licensed data sets provided to analysts, and on-site access.

Data dissemination is the finished product and comes in many forms, from reports, to machine-readable datasets, to dashboards and websites. There are plenty of technologies that can support dissemination. It is important to understand how the technology manages access to the finished product, disseminates data, and minimizes the risk of redisclosure. There are multiple open source and commercial options to support dissemination in an accessible and secure way. **Tableau**, **ESRI's ArcGIS**, **PowerBI**, **Qlik**, **Cognos**, **Business Objects**, and **Oracle OBIEE** are some of the most common tools leveraged for data visualization.

When thinking about data access and dissemination, ask these questions:

1. How are integrated data accessed by analysts?

2. Who gets access to the data infrastructure, and by what means? Are there levels of access for identified and de-identified information, or for internal and external users?

3. Are there aspects of the collection and processing pipeline which must be taken into account to ensure statistical rigor? For example, are there ways in which quality control might systematically exclude data from certain individuals or organizations, or ways in which subtleties in the entity definitions might confound an analyst without deep familiarity with the domain?

4. How are integrated data products disseminated?

5. How do your data partners share their data? Via spreadsheets and shared drives, via API, or via some other mechanism?

6. What software solution(s) work best with the infrastructure? For each solution, is it modular, and does it have a documented API?

7. How are infrastructure changes made? If there is a vendor, do they need to be involved in every change, and if so, what are the expectations around timeline and cost?

Technology For Civic Data Integration: Five-Year Outlook

- Staff skills should move up in the value chain, especially as the technologies evolve. Technology will facilitate people to provide higher value deliverables; **how we integrate data will no longer be the question**, but rather how we ensure that the analysis and models are affecting practice and driving change.

- The field has come a long way in the past five years, particularly around transparency and consistency. In order to conduct multi-site inquiry and continue to improve data models, all sites must emphasize the **development of metadata**; not only as a data operation but to contribute to the field.

- Experts and practitioners will be thinking about **blockchain**, consent management and identity management in **how we collect data on people**. We will need to be mindful of artificial intelligence and machine learning to ensure that we are careful in training data sets while focusing on ethical use and data governance practices.

- Infrastructure should provide more than just integration support, it **should provide transparency and documentation about how solutions were developed**, including confidence levels for results.

- The data science "gap" will diminish as schools incorporate analytic methods and skills into more undergraduate and professional graduate programs. An **ongoing bottleneck will be legal and transactional issues** such as privacy, domain use, ethics, international considerations, and inclusive engagement.

Conclusion

The purpose of this report is to describe key considerations in building and sustaining IDS and the various technology approaches for overcoming challenges in data integration. Sharing, integrating, and using administrative data to inform policy and practice is technically complex, and involves navigating myriad constraints around staffing, security, management, processing, storing, and sharing data.

- First and foremost, it is essential to ensure you have the staffing expertise and data management processes in place to effectively manage the infrastructure and to minimize the risks associated with integrating administrative data.

- Evaluating data management needs requires close examination of how to drive standardization as much as possible, upstream and downstream, and simultaneously plan for an agile and flexible data intake and normalization or indexing process. Pursuing both in parallel will reduce the dependence on specific vendors, reduce ongoing maintenance and upgrade costs, and maintain an ability to link new data streams as they become relevant.

- The earliest decisions to make regarding technology solutions for IDS are "how will the source data get into the IDS" and "how will the IDS protect the data?" While there are many ways to upload data, this paper categorizes them as either manual (an individual uploads the data) or automatic (there is a connection between source and integration system). In many infrastructures, a hybrid approach that allows for some manual uploads and some automated connections is implemented. When integrating data, organizations have to ensure that data is protected and secured from the moment it enters the IDS to when it is disseminated. This means ensuring that the integrity of the data and individual privacy is legally, technically, procedurally, and physically maintained throughout the process.

- There are plenty of options for storing data, related to both how and where the data will be stored. The decision to store data on-site, in a data center, or leverage the many cloud-based solutions available to the IDS will most likely require a warehousing solution to properly store and make data accessible. What you choose is largely dependent upon the backup needs of the organization as well as the amounts of data that will be stored and accessed via the system.

- How the IDS supports linking, access, and data dissemination is often what makes or breaks adoption of the IDS. There are two key methods for data linkage: deterministic matching and probabilistic matching. Data dissemination is the finished product and comes in many forms, from reports, to machine-readable datasets, to dashboards and websites. There are plenty of technologies that can support dissemination and it is important to understand prior to procurement how the technology manages access to the finished product, how it will be disseminated, and how risks of redisclosure are minimized.

Appendix 1: Key Terms

Advanced Encryption Standard (AES)—The Advanced Encryption Standard (AES) is a specification for the encryption of electronic data established by the U.S. National Institute of Standards and Technology (NIST) in 2001.

Application Programming Interfaces (API)—In computer programming, an application programming interface (API) is a set of subroutine definitions, communication protocols, and tools for building software. In general terms, it is a set of clearly defined methods of communication between various components. A good API makes it easier to develop a computer program by providing all the building blocks, which are then put together by the programmer. An API may be for a web-based system, operating system, database system, computer hardware, or software library. An API specification can take many forms, but often includes specifications for routines, data structures, object classes, variables, or remote calls. POSIX, Windows API and ASPI are examples of different forms of APIs. Documentation for the API is usually provided to facilitate usage and implementation.

Data Aggregation—is the compiling of information from databases with intent to prepare combined datasets for data processing.

Data Encryption Standard (DES)—The Data Encryption Standard (DES) is a symmetric-key algorithm for the encryption of electronic data.

Data Masking/ Data Obfuscation—is the process of hiding original data with random characters or data.

Data Model—A data model is an abstract model that organizes elements of data and standardizes how they relate to one another and to properties of the real world entities. Data models describe the structure, manipulation and integrity aspects of the data stored in data management systems such as relational databases.

Data Vault—is a hybrid data modeling methodology providing historical data representation from multiple sources designed to be resilient to environmental changes.

Dimensional Data—In a data warehouse, dimensions provide structured labeling information to otherwise unordered numeric measures. The dimension is a data set composed of individual, non-overlapping data elements. The primary functions of dimensions are threefold: to provide filtering, grouping and labelling.

Elliptical—Curve Cryptography (ECC)- Elliptic-curve cryptography (ECC) is an approach to public-key cryptography based on the algebraic structure of elliptic curves over finite fields. Elliptic curves are applicable for key agreement, digital signatures, pseudo-random generators and other tasks. Indirectly, they can be used for encryption by combining the key agreement with a symmetric encryption scheme.

Extraction/Transformation/Loading (ETL)—'Extract, Transform, and Load' is a common 3-step process for blending data from multiple sources to build a data warehouse. Data are taken from a source (extract), converted into the appropriate format for analysis (transformed), and stored (loaded) into the data storage system.

APPENDIX 1: KEY TERMS

Family Educational Rights and Privacy Act (FERPA)—is a federal law that protects the privacy of student education records. The law applies to all schools that receive funds under an applicable program of the U.S. Department of Education.

FedRamp—The Federal Risk and Authorization Management Program (FedRAMP) is a government-wide program that provides a standardized approach to security assessment, authorization, and continuous monitoring for cloud products and services.

Health Insurance Portability and Accountability Act (HIPAA)—HIPAA (Health Insurance Portability and Accountability Act of 1996) is United States legislation that provides data privacy and security provisions for safeguarding medical information. The law has emerged into greater prominence in recent years with the proliferation of health data breaches caused by cyberattacks and ransomware attacks on health insurers and providers.

International Classification of Diseases (ICD)—The International Classification of Diseases (ICD) is the international "standard diagnostic tool for epidemiology, health management and clinical purposes." Its full official name is International Statistical Classification of Diseases and Related Health Problems.

LOINC—LOINC is a common language (set of identifiers, names, and codes) for identifying health measurements, observations, and documents.

Nonrelational (or NoSQL) Databases—is any database that does not follow the relational model provided by traditional relational database management systems. This category of databases, also referred to as NoSQL databases, has seen steady adoption growth in recent years with the rise of Big Data applications. Nonrelational databases have grown in popularity because they were designed to overcome the limitations of relational databases in dealing with Big Data demands.

Payment Card Industry Data Security Standard (PCI DSS)—is an information security standard for organizations that handle branded credit cards from the major card schemes. The PCI Standard is mandated by the card brands and administered by the Payment Card Industry Security Standards Council. The standard was created to increase controls around cardholder data to reduce credit card fraud. Validation of compliance is performed annually, either by an external Qualified Security Assessor (QSA) or by a firm specific Internal Security Assessor (ISA) that creates a Report on Compliance for organizations handling large volumes of transactions, or by Self-Assessment Questionnaire (SAQ) for companies handling smaller volumes.

Role-Based Access Controls (RBAC)—is a method of regulating access to computer or network resources based on the roles of individual users within an enterprise. In this context, access is the ability of an individual user to perform a specific task, such as view, create, or modify a file.

RSA—RSA is an algorithm used by modern computers to encrypt and decrypt messages. It is an asymmetric cryptographic algorithm. Asymmetric means that there are two different keys. This is also called public key cryptography, because one of the keys can be given to anyone.

SNOMED—is a systematically organized computer processable collection of medical terms providing codes, terms, synonyms and definitions used in clinical documentation and reporting.

Standardization—Data standardization is the process of converting variegated data formats into one format. IDS utilizes multiple databases from a variety of service provision settings. For example, even within research on one administrative domain, say healthcare, a study might integrate ambulance records, insurance claims, provider reimbursement, and pharmacy data. The same concepts may be represented by different formats or constructs across these datasets. Data standardization is key to ensuring data quality in research. An integrated project lacking standardization would compromise the accuracy of results.

Statement on Standards for Attestation Engagements no. 16 (SSAE 16)—is an auditing standard for service organizations.

Structured Data—Data that have be formatted and stored in a repository (e.g., a database) making them easy to process and analyze.

Unstructured Data—Data that have not been organized into a pre-defined data model. Unstructured data lack a format, meaning that it is difficult to access and generate actionable information from these data. Unstructured data are often text heavy, but can also include numbers as well. They include file formats like emails, videos, and photos that don't fit neatly in a database.

APPENDIX 2: EXAMPLES OF JOB DESCRIPTIONS

Please note the examples provided have been scrubbed of identifying information and should be considered a jumping off point for your own endeavors. Neither MetroLab, NNIP nor AISP endorse any of the samples below but would like to thank the various agencies and institutes that let us use this content.

TABLE OF CONTENTS LINKS

Sample Position: Senior Information Architect

Definition: Under general supervision, to analyze and translate business information and technical requirements into an information and data architectural blueprint for the IT system.

Distinguishing Characteristics

Positions in the Information Architect series are responsible for improving business performance or cost optimization through enterprise information solutions, such as master data management, metadata management, business intelligence, data catalogs, data warehouses, data interoperability, analytics, data integration and related information infrastructure components. Incumbents are principally focused on the high-level information repositories and information flows that will meet the requirements of the business strategies and business processes. Information Architects provide guidance and training to Reporting/Informatics Analyst and Data Engineers. They work collaboratively with business and IT stakeholders to define Information Architecture Standards as well as develop roadmaps for achieving and maintaining standards. Information Architects help curate and enhance the quality and usability of agency data assets.

The Associate Information Architect classification is distinguished from the Information Architect classification in that the Associate Information Architect applies a comprehensive knowledge of a particular field of specialization to assignments where the content objectives are clearly defined, content and information is simple and targeted to a small audience, and the content is served on a single platform. The Associate Information Architect is involved in one or more projects; serves as a team member or internal consultant; and functions with autonomy, requiring occasional guidance from project leadership.

The Information Architect classification is distinguished from the Senior Information Architect classification in that the Information Architect applies an advanced knowledge of a particular field of specialization to assignments where the projects are larger and more important or there are multiple projects, the content objectives are somewhat defined, content and information is collected and provided in different media, the target audience is mixed with diverse needs and understanding, and the content is designed for multiple platforms such as web, mobile, smart devices etc. The Information Architect serves as a team member, internal consultant or team lead; analyzes unique issues/problems that are without precedent and/or structure; and functions with a high degree of autonomy, with occasional guidance from senior leadership.

The Senior Information Architect classification is distinguished from the Information Architect classification in that the Senior Information Architect applies an advanced knowledge of a particular field of specialization to projects with mission critical complexity. Mission critical projects have vague content objectives, the data and documents are

complex and in multiple formats, there is a highly diverse audience with different needs and understanding, and the content is designed for multiple platforms such as web, mobile, smart devices etc. The projects are IT enterprise-wide. The Senior Information Architect regularly leads projects of critical importance to the agency; these projects carry substantial consequences of success or failure; leads or directs projects and provides technical leadership and consultation; and functions with the highest degree of autonomy with occasional guidance from executive leadership.

Typical Tasks

1. Confirms regulatory issues are considered surrounding information assets (such as consumer privacy laws, data retention policies, outsourced data and specific industry guidelines are adhered to;
2. Monitors data quality, and analytical systems usage, and takes corrective and enhancement actions as necessary
3. Leads the IT enterprise-wide technical and business discussions relative to future enterprise direction;
4. Develops IT strategy and aligns the strategy with organizational goals;
5. Establishes the scope of information architecture activities, including external (industry) collaboration areas;
6. Oversees the implementation of a discipline and approach for information assets;
7. Consults on the identification of information that the organization produces and consumes;
8. Consolidates a business view of data across the education, research and administrative functions;
9. Reviews and approves data model designs with a focus on those areas influencing interoperability, analytics including compliance with the portfolio information model;
10. Reviews and/or analyzes and develops architectural requirements for the IT enterprise-wide level;
11. Approves modification of the information architecture to accommodate project needs;
12. Leads the development and execution of a communication and education plan for the IT system architecture; responsible for the overall training and development of the architect family staff;
13. Collaborates in defining solution architecture for Master Data Management (MDM);
14. Reviews and approves the strategic requirements, principles, models and designs that strengthen the network effect of sharing such information across the IT system;
15. Collaborates in the definition and implementation of the enterprise disaster recovery approach and plan for the recovery of the IT service portfolio and environment, to meet the need for immediate business continuity and resumption;
16. Provides strategic consultation to clients and IT teams; advises on options, risks, costs versus benefits, system impacts, and technology priorities;
17. Develops IT enterprise-wide policies, standards, guidelines and procedures pertaining to information architecture, sharing and integration;
18. Collaborates with other IT teams in the management of technical change, verification of testing results, and monitoring of technical standards compliance and deployment;
19. Reviews and approves the organizational requirements for the resources, structures and cultural changes necessary to support initiatives;
20. Performs other related duties as required.

Training and Experience Note:

Bachelor's degree in Computer Science, Information Systems, Computer Engineering, System Analysis or a related field from an accredited college and fifteen (15) years of IT and business/industry work experience including data modeling, data integration, metadata management, master data management, business intelligence and analytics, or establishment of data architecture.

Knowledge of:

1. Principles of leadership;
2. External (industry) collaboration areas;
3. Business view of data across the education, research and administrative functions;
4. Principles of training and staff development;
5. Regulatory issues related to information assets;
6. IT system strategy;
7. Information architecture standards and methodologies;
8. Strategic requirements, principles, models and designs related to information architecture;
9. Information structure and features, functionality, and user-interface design;
10. Data valuation process for data objects;
11. Well-integrated development activities;
12. Data mappings of application interfaces and data models;
13. Information architecture models (e.g., conceptual, logical, canonical);
14. Roadmaps analysis, design, and implementation;
15. Enterprise disaster recovery and planning;
16. Organization requirements for the resources, structures and cultural changes necessary to support initiatives;
17. Applicable policies, standards, guidelines and procedures pertaining to information architecture;
18. Technical change, verification of testing results, and monitoring of technical standards compliance and deployment;
19. More than one design method and at least one discipline that is considered to be at the level of a subject matter expert;
20. Business processes in the work environment;
21. Guidelines and standards for cataloging and maintaining data repositories;
22. Single source data strategy;
23. Data design and transformation procedures;
24. Data models, associated metadata models, common business vocabulary, ontologies and taxonomies;
25. Consumer privacy laws, data retention policies, outsourced data and specific industry guidelines;
26. Mapping and audit-ability of information assets;
27. Site maps, use cases, business process flows, and navigational flows;
28. Conflict resolutions among models, ensuring that data models are consistent with the enterprise model (e.g.: entity names, relationships and definitions;
29. High level modeling of key program processes and key data entities;
30. Methods to align architectural strategy with business requirements;
31. Analysis, design, and development of a roadmap and implementation plan based upon a current versus future state in a cohesive architecture viewpoint;
32. Specialized architecture governance processes;
33. Post-implementation continuous-improvement efforts to enhance performance and provide increased functionality.

Ability to:

1. Lead technical and business discussions;
2. Align strategy with organizational goals;
3. Consolidate business view of data across the education, research and administrative functions;
4. Review and approves data model designs;
5. Lead the development and execution of a communication and education plans.
6. Coordinate development activities to ensure consistency and integration;
7. Create information architecture models (e.g. conceptual, logical, canonical);
8. Analyze, design, and develop roadmaps and implementation plans;
9. Define and implement disaster recovery approach plans;
10. Implement data design and transformation procedures;
11. Develop usable information architecture;
12. Adhere to applicable regulations surrounding information assets;
13. Create a framework for representing the data elements;
14. Resolve conflicts between data models;
15. Consult and collaborate with multiple IT teams and business leaders;
16. Write effective policies, standards, guidelines and procedures;
17. Gather and analyzes data at the project/program level;
18. Assess business needs and establish business priorities.

Sample Position: Data Engineer

Definition: Under general supervision, to lead staff in the building and maintaining data processing systems and pipelines that combine core data sources into accessible structures (e.g. data warehouses, data marts, data lakes, etc.) to support reporting and analytical systems; to collect, store, and clean data to make it available to Data Analysts, Data Scientists, or Business Intelligence Analysts who can easily query the data.

Distinguishing Characteristics:

Positions in the Data Engineer series are involved in designing, coding, testing and implementing data extract, transform and load (ETL) procedures and work closely with Information Architects, Data Analysts, and Data Scientists to determine what data management systems are appropriate and which data are needed for analysis. The Data Engineer must have strong software engineering skills and database management systems knowledge to develop a clear understanding of the business needs and incorporate these needs into technical solutions

The Associate Data Engineer classification is distinguished from the Data Engineer classification in that the Associate Data Engineer works on one or more projects as a team member; designs, normalizes, develops and implements low to medium complex database management systems, where the requirements are well defined or familiar, and the number of affected agencies/departments is typically five or less. They monitor the status of data processing jobs and report out on errors and inconsistencies.

The Data Engineer classification is distinguished from the Associate Data Engineer in that the Data Engineer classification works on multiple projects as a team member; designs, normalizes, develops and implements medium to highly complex database management systems, where the requirements and objectives may be unknown at the start, and the number of affected agencies/departments is often greater than five. They work with technology and business stakeholders to triage and help correct data processing issues.

The Senior Data Engineer classification is distinguished from the Data Engineer classification in that the Senior Data Engineer classification works on multiple projects as a technical lead; designs, normalizes, develops and implements highly complex and/or enterprise level database management systems which may involve complex integrations and new technologies; reviews and approves database designs to ensure integrity, availability and confidentiality of data processing.

Typical Tasks

1. Leads data teams to align data pipeline design with overall solution design;
2. Works closely with Data Analysts and Information Architects to assure all systems are in line with strategy and overall solution design;
3. Defines and implements data hygiene, normalization, and data enrichment processes;
4. Lead post-implementation continuous improvement efforts in enhancing performance and helps developers to create and or performance tune database queries;
5. Collaborates in the development of data standards, policies and procedures;
6. Designs and creates data processing systems that combine core data sources into data repositories that support reporting and analytical systems;
7. Enhances existing database systems to increase operating efficiency or adapt to newer requirements;
8. Analyzes performance and capacity and advises any necessary infrastructure changes;
9. Recommends and implements data reliability, efficiency and quality improvements;
10. Researches opportunities for data acquisition and new uses for existing data;
11. Contributes to work plan timelines and manages workflows to meet project timeframes;
12. Designs and creates database management systems that combine core data sources into data warehouses to support reporting and analytical systems;
13. Performs data conversions, imports and exports of data within and between internal and external software systems;
14. Creates data transformation and automation processes using a variety of technologies (e.g. Scripting, PowerShell, ETL, SQL stored procedures, etc.) to support business systems and data flows;
15. Maintains the quality of data repositories by adding, modifying, and deleting data;
16. Troubleshoots data processing tools, systems, and software;
17. Identifies and resolves production and/or application development problems that relate to the use of the database management systems;
18. Monitors performance and capability according to defined service level agreements;
19. Develops and maintains a query library to support recurring data requests;
20. Documents new and existing models, solutions, and implementations such as data mapping, technical specifications, production support, data dictionaries, test cases, etc.;
21. Maintains the data dictionary and other related data processing metadata;
22. May be assigned as a Disaster Service Worker, as required;
23. Performs other related duties, as required.

Training and Experience Note:

Bachelor's degree from an accredited college in Science, Technology, Engineering or Math and nine (9) years of experience working in information systems development, focused on processing large volumes of data feeds or work experience as a Data Base Administrator (DBA) on Data Base Management System design and support working in query languages and Extract, Transform and Load (ETL) technologies and tools.

Knowledge of:

1. Methods to align database design with overall solution design;
2. Principles of leadership and teamwork;
3. Data hygiene, normalization, and data enrichment processes;
4. Post-implementation continuous improvement efforts in enhancing performance;
5. Advanced skills in analytical data processing tools and techniques;
6. Information systems development across the IT lifecycle, with a focus on systems to perform high volume and velocity data processing on disparate data types/formats;
7. Query languages and various ETL technologies and tools (e.g. SQL, Informatica, SSIS, Data Stage, Alteryx, Denodo, Snowflake, etc.);
8. Relational and NoSQL database technologies;
9. Database management software and defining hardware requirements;
10. Software components (e.g., specialized UDFs) and analytics applications;
11. Data/information quality metrics;
12. Troubleshooting techniques of data load or reconciliation;
13. Methods used in vendor and software evaluations;
14. End-user training and training materials;
15. Relational database theory, structure, principles, and practices, database normalization concepts, data modeling and performance tuning;
16. Software development methodology and release processes;
17. Data analytics and business intelligence tools.
18. Data conversions, imports and exports of data within and between internal and external software systems;
19. Processes to ensure data integrity and standardization;
20. Data quality standards;
21. Data analytics tools, systems, and software troubleshooting;
22. Resolutions for production and/or application development problems that relate to the use of the database management systems;
23. Data management standards, policies and procedures;
24. Disaster recovery procedures.

Ability to:

1. Lead data teams to align database design with overall solution design;
2. Work with other IT staff to assure all systems are in line with strategy and overall solution design;
3. Define and implement data hygiene, normalization, and data enrichment processes;
4. Lead post-implementation continuous improvement efforts in enhancing performance;
5. Collaborate in the development of data standards, policies and procedures;
6. Design and create database management systems;
7. Recommend and implement data reliability, efficiency and quality improvements;
8. Research opportunities for data acquisition and new uses for existing data;
9. Enhance existing database systems;
10. Perform data conversions, imports and exports of data within and between internal and external software systems;
11. Create data transformation processes (ETL, SQL stored procedures, etc.) to support business systems and operational data flows;
12. Maintain the quality of data repositories;
13. Troubleshoot data analytic tools, systems, and software;
14. Identify and resolve production and/or application development problems;
15. Analyze performance and capacity and advise on infrastructure changes;
16. Develop and maintain a query library to support recurring data requests;
17. Document new and existing models, solutions, and implementations;
18. Maintains the data dictionary.

Sample Position: Senior Data Analyst

Definition: Under general supervision, to identify, capture, manage, and analyze the agency's data to provide data analysis including the conceptualization, modeling and presentation of data.

Distinguishing Characteristics:

Positions in the Data Analyst series perform a variety of tasks that contribute to better decision-making and planning. Incumbents provide data analysis including the conceptualization, modeling and presentation of data. Additional responsibilities include supporting information governance along with Information Architects and departmental data stewards.

The Associate Data Analyst classification is distinguished from the Data Analyst classification in that the Associate Data Analyst applies fundamental concepts, practices and procedures to complete assignments of low complexity, where the requirements/outcomes are well defined or familiar. The Associate Data Analyst works on one or more data and analytics projects and works as a team member.

The Data Analyst classification is distinguished from the Associate Data Analyst classification in that the Data Analyst applies the fundamental concepts, practices and procedures to complete moderately complex assignments, where the requirements/outcomes are ill-defined or partially defined. The Data Analyst works on multiple data and analytics projects and works as a team member. The Data Analyst provides training and education to business users on how to find, understand and combine data for advanced analytical purposes.

The Senior Data Analyst classification is distinguished from the Data Analyst classification in that the Senior Data Analyst applies a broad knowledge of principles, practices, and procedures to complete highly complex assignments, where the requirements and outcomes are unfamiliar or unknown. The Senior Data Analyst works on multiple data and analytics projects and program execution; defines the requirements for analysis within multiple business area or enterprise-wide; and works as a team member, a technical project lead, or a data analyst advisor.

Typical Tasks

1. Works with Senior Leadership to determine data analytic organizational goals;
2. Works with Data Scientists to identify and reveal trends, patterns, and relationships in data to provide guidance for Senior Leadership decision making processes;
3. Provides expertise in determining the right type of analytics or statistics to apply to data sets collected and utilized by the customer;
4. Provides input on data governance policy creation;
5. Collaborates with data governance board members to receive and understand policy directives;
6. Develops standards and best practice procedures for data validation, analysis and presentation;
7. Creates and implements techniques for data collection, accuracy and management;
8. Provides data analysis coaching to team members;
9. Queries and profiles data to determine conformance to standards and fit for purpose;
10. Maintains up-to-date knowledge of regulatory requirements and industry best practices as it relates to data stewardship;
11. Trains users on data structure and meaning within analytics platforms to enable self-service reporting;
12. Triages data-related issues around quality and analytical production;
13. Pinpoints trends, correlations and patterns in complicated data sets;
14. Documents and maintains all data management procedures and specifications;
15. Provides guidance to data stewards in the range of stewardship activities and best practices;
16. Identifies opportunities to increase data quality, including automation and data entry procedural training;
17. Catalogs existing data sources;
18. Maintains relational databases and data systems;
19. Works with open datasets, including populating and extracting data from internal and public open datasets;
20. Designs reporting dashboards to translate data to an understandable illustrative format;
21. Creates data visualizations and reports using presentation layer tools;
22. Extracts and manages multiple sets of data from various databases and sources, bringing together data from several channels for comparative analysis;
23. Mines and analyzes data using standard statistical tools and techniques;
24. Reviews and validates data for accuracy; performs data cleansing as needed;
25. Documents and maintains record of all data management activities;
26. May be assigned as a Disaster Service Worker, as required;
27. Performs other related duties, as required.

Training and Experience Note: Bachelor's degree from an accredited college in Computer Science, Information Technology, Mathematics, Statistics or Engineering and six (6) years of work experience directly related to data analysis or statistical analysis.

Knowledge of:

1. Advanced analytics and statistical models to identify trends, patterns and relationships;
2. Techniques for data collection and management;
3. Principles of leadership and team coaching;
4. Government programs regulations, requirements and best practices;
5. Data stewardship regulatory requirements and industry best practices;
6. Trends, correlations and patterns in data sets;
7. Methods to increase data quality;
8. Spreadsheets and databases, such as MS Access;
9. Relational databases, NoSQL databases, data extraction, querying, and scripting;
10. Scripting and coding to automate and monitor data management processes;
11. Data visualizations and reporting tools to present information to a variety of audiences using presentation layer tools;
12. Extracting and managing multiple sets of data from various databases and sources, bringing together data from several channels for comparative analysis;
13. Initial data exploration steps (binning, pivoting, summarizing and finding correlations, for example);
14. Standard statistical tools and techniques.

Ability to:

1. Work collaboratively with all levels of management, customers and staff;
2. Provide effective leadership for team members;
3. Triage code problems;
4. Implement statistical modes to identify trends, correlations and patterns in data sets;
5. Increase data quality;
6. Catalog data sources;
7. Maintain, design and create relational databases and data systems;
8. Design reporting dashboards;
9. Use presentations tools;
10. Manage multiple sets of data;
11. Mines and analyze data;
12. Perform data cleansing;
13. Document and maintain records.

Sample Position: Lead Systems Administrator

Position Summary: Implements, installs, monitors, and documents city-wide IT systems infrastructure. This position serves as a lead for IT systems administration staff, while receiving supervision and management from the Tech Infrastructure Manager.

Work Experience: The application must clearly show three (3) years of full-time systems administration experience which includes: performing systems maintenance, implementation and support.

Education/Training: The application must clearly show a Bachelor's Degree from a fully accredited institution in Computer Science, Information Science, Business Administration, or a related field.

Equivalency: Education/training and/or work experience may be substituted on a year for year basis if the application clearly shows the required number of years to meet the Total Qualifying Requirement for this position (based on the sum of work experience and education/training listed above).

Position Duties

1. Provides support and maintenance city-wide for servers running various Operating Systems; researches and tests new software products as they relate to server technology.
2. Provides escalated support in the Installation, configuration, testing and support of client computer hardware, software (OS and Application), printers, communication devices, graphics and storage peripherals.
3. Recommends purchase and/or modification to existing computers and peripheral and network software and operating systems.
4. Plans, implements and supports data transfer concepts with internal and external entities by various technologies
5. Designs, develops, and maintains the City's disaster recovery co-location facility.
6. Oversees and validates the City's backup and recovery model to maximize retention based on departmental policy.
7. Supports City's IT infrastructure (i.e. Physical Server Hosts, Storage Infrastructure and various interconnect technologies, Cloud based resource implementations).
8. Participates in installation and physical maintenance of hardware systems
9. Participates in project-oriented tasks, including team-scale project leadership.
10. Oversees the team's help desk ticket assignments and flow to ensure expectations for quality, speed and work balance are met.
11. Reviews agreed upon team KPIs to assess System Administration team performance.
12. Works with Technology Infrastructure Manager and departmental leadership to identify and implement new process improvements that will improve the level of service provided by the System Administration team.
13. Serves as an escalation point for other members of the System Administration team who need assistance resolving client issues.
14. Provides support for the team's professional and technical skills development needs.
15. Works with Technology Infrastructure Manager to balance workload among team members as new projects arise.

16. Works with Technology Infrastructure Manager to develop and update content for team knowledge base, including standard operating procedures.

17. Acts as a liaison with consultants, vendors, outside agencies, City Departments, etc., on assigned projects.

18. Manages area of responsibility in strict accordance with applicable laws, regulations and established policies, including union agreements, to ensure fair and standardized treatment of employees.

19.Plans, develops, directs, coordinates, organizes and controls the resources under jurisdiction toward the effective and efficient attainment of organizational and program goals, including public service and employee safety.

20. Ensures compliance with vendor licensing laws and regulations related to software applications.

21. Ensures adherence to established departmental computer system and network standards, policies and procedures.

22. Assists in preparing bidding and purchase specifications.

23. Maintains records, both manually and by computer, and prepares accurate reports, correspondence, etc.

24. Performs activities and functions of related lower-level personnel.

25. Performs other related tasks and duties as assigned or required.

Knowledge, Skills, and Abilities

Extensive knowledge of:

1. Virtualization Technologies including virtual servers, desktops, networking and storage concepts.
2. Disaster recovery co-location and recovery planning
3. Windows File share and Permission Management concepts.
4. Scripting methodologies.
5. VMware, ESX, Horizon View, NSX, UCS, HP 3Par, PowerShell, and F5 technologies.
6. Server configuration, design, maintenance, performance and capacity planning.
7. Understanding of networking concepts such as DNS, DHCP, Email, HTTP, SSL, OSI Model, and TCP/IP protocols and applications.
8. Load balancing concepts including resource pooling and resource masking.
9. Platform / Software as a Service concepts, including hybrid implementations.
10. Client-type computer hardware and peripherals.
11. Troubleshooting methodologies.

Ability to:

1. Diagnose and troubleshoot in the areas of systems applications, operating hardware, operating software and networking.
2. Provide effective technical assistance and training to users.
3. Supervise, monitor, evaluate and direct the work of assigned personnel.
4. Effectively ensure the security and integrity of computer systems and networks including the Internet and intranet.
5. Establish and maintain records, both manually and by computer, and prepare accurate reports, correspondence, etc.
6. Communicate effectively, both orally and in writing.
7. Work outside regular hours and ability to be on call on some weeknights and weekends.
8. Establish and maintain effective working relationships with supervisors, associates and outside agencies.
9. Use analytical skills to thoroughly dissect a problem and present viable solutions to management for discussion.

Sample Position: Integrated Data System Director

Position Overview:

Organization is seeking an Integrated Data System (IDS) Director to work with our Senior Director of Data and Policy Partnerships and our Executive Director to direct our organization's efforts to create an integrated data system in collaboration with partner organizations that will contribute administrative data to the IDS. The IDS Director is a position that requires a unique combination of skills. The IDS is in its infancy, so a successful applicant will enjoy public and/ or nonprofit entrepreneurial endeavors and welcome the challenges of building new programs across multiple systems partners.

The IDS Director will need excellent people skills as s/he will work to cultivate relationships with data-contributing organizations, researchers, and community stakeholders. The ideal candidate will possess well-developed leadership skills, be proactive yet patient, and be able to simultaneously manage multiple aspects of a complex, collaborative project. The IDS Director should possess a firm understanding of the issues of security, privacy, ethics, and public trust inherent to data sharing and data analysis of social issues. Ideally, the IDS Director will be familiar with the local context and public institutions and s/he will have experience with the legal and technical aspects of data-sharing and data use partnerships. The IDS Director must apply a race equity and inclusion lens to all aspects of IDS creation from partner agreements to data linkage to community outreach to IDS implementation.

IDS Director Essential Duties and Responsibilities

IDS Development (Immediate):

1. Work with the Senior Director of Data and Policy Partnerships and the Executive Director to cultivate relationships with staff-members of current and prospective data contributing agencies.
2. Work with agencies to identify areas in which integrated data would help build agency capacity in diverse ways such as improving intelligence available for program design, service provision, or other essential functions of the organization.
3. Build coalitions of data practitioners at key agencies and data stakeholders in the community to generate understanding of, support for, and participation and use of the IDS.
4. Work with the IDS site team to develop a data sharing landscape and a data inventory for location. Identify/ update landscape of relevant data that would ideally be integrated into an IDS.
5. Work with agency partners to identify data fields that would be valuable for linking to data of other agencies and identify potential impediments to linkage and standardization of fields.
6. Help Site Team develop and implement a community outreach strategy around integrated data needs and uses.
7. Identify opportunities to enrich public understanding of cradle-to-career outcomes through inter-agency linked data.
8. Work to convene events and meetings that promote the cause of integrated data through education, outreach, and applied examples.
9. Assist with development of foundational documents including IDS governance policies and processes and communications documents.

IDS Implementation (Future):

1. Work with Senior Director of Data and Policy Partnerships, data partners, and technical assistance organizations to hire and manage contractors performing services related to data linking, data auditing, and security architecture.

2. Oversee data linking process including standardization and documentation.

3. Manage the administration of the Integrated Data System (IDS). Oversee day-to-day management, use, and security of the Integrated Data System (IDS).

4. Ensure compliance with IDS governance policies and processes.

5. Work with agency and organizational representatives on implementation teams to facilitate IDS use as approved by Governing Board and in line with IDS policies.

5. Work with Director of Data Partnerships and IDS Site Team to develop and maintain materials on IDS use policies and procedures as well as IDS communications materials for engaging public sector and service provider practitioners, research partners, and community stakeholders

6. Support implementation of collaborative research projects as approved by IDS Governing Board with an emphasis on participatory research and community engagement.

7. Oversee reporting out of IDS research results and convening around findings in accordance with governance policies and requirements.

8. Supervise periodic data auditing process.

Preferred Education, Knowledge, Skills, and Experience

1. Minimum of a Master's Degree in field related to social science research including but not limited to applied data science, public health, economics, sociology, public policy, political science, psychology, public administration, education, or another related field.

2. Interpersonal skills with ability to cultivate relationships with perspective and patience and to conduct interactions with sensitivity and diplomacy.

3. Leadership skills: ability build consensus, create coalitions, and manage multiple project elements while adhering to IDS guiding principles pertaining to ethics, accountability, transparency, and public trust and applying a race equity and inclusion lens to all aspects.

4. Advanced training and experience in quantitative methodology as well as qualitative methods. Experience with applied data analysis.

5. Familiarity with location public agencies institutions including public education, social services, public health agencies among others.

6. Understanding of social, racial, and economic context of location. Ability to apply race equity lens to data issues ranging from matching methods to results interpretation.

7. Commitment to collective impact and collaborative process

8. Previous experience integrating data; familiarity with technical issues (e.g. cleaning, merging, standardizing, documenting, and securing) as well as legal issues (e.g. MOU development, data use agreements.)

9. Appreciation for importance of and knowledge of advanced measures for ensuring integrity and transparency in all aspects of integrated data linkage and use and ability to create and enforce protocols for adhering to and documenting those processes.

10. Previous experience in nonprofit or public-sector environment or a combination thereof.

11. Excellent communications skills including oral and written communication as well as public presentations.

Sample Position: Director of DevOps

Basic Function: Oversee the work of the DevOps team. Ensure that the web applications created by the DevOps team meet the needs of organization and our external partners. Manage the Senior Information Technologists and help guide the planning and software development practice. Oversee the importation and integration of partner data into the DataHub.

Essential Duties and Responsibilities:

Oversee development efforts by Senior Information Technologists. Work with Senior Information Technologists to prioritize work for downtime. Identify professional development opportunities for team. Work with DevOps team to improve process and programming practice. Recruit new DevOps Senior Information Technologists and DevOps Interns as needed. Participate in the onboarding and orientation of new staff.

Collaborate with other members of organization's senior staff to establish priorities, timelines, deliverables, and product specifications. Engage in strategic planning around new technology, data security, contingency planning, and team capacity. Work with analysts and partners to explain data linkage methodology. Consult with partners on potential new projects. Provide estimates of time, effort, and dependencies. Provide support or training where necessary for analysts and partners using organization products. Advise on technology and data strategy where necessary.

Oversee the process for importing and linking data in the DataHub. Work with organization to ensure data is secure, and recommend policies related to data security. Work with organization Director to secure data-sharing agreements. Coordinate with analysts, partners, and Senior Information Technologists to create new data models and receive data to be included in the Data Hub. Work with DevOps team to test data imports and linkage, ensuring data integrity and quality.

Find opportunities to improve DataHub's linkage quality and performance. Identify opportunities for routinizing common data management tasks. Manage software configurations and access controls. Work with IT to formulate and implement data backup and recovery procedures. Collaborate with IT to maintain and update the business continuity plan and other critical infrastructure documentation. Research and implement programming best practices and new technologies related to data management.

Act as product owner for organization's core products. Evaluate new feature requests and opportunities and maintain a backlog of user stories for each product. Prioritize programming work based on need, value, and budget. Work with team on quality assurance and testing of new features. Work with external vendors as needed. Coordinate with Senior Information Technologists on upgrade path for underlying technologies. Work on defining new products when applicable, such as data dashboards for partner agencies.

Licenses, Tools and Equipment: Web applications, software, data models, personal computers, printers, fax machines, scanners and word processing.

Required Qualifications:

1. Bachelor's degree or higher in Computer Science or related field;
2. Minimum five years' to include Python, SQL, and Django; Demonstrated project management experience;
3. Demonstrated product management experience; Demonstrated knowledge of relational databases, SQL, and PostgreSQL;
4. Demonstrated knowledge of common data structures and algorithms;
5. Demonstrated strong interpersonal and verbal communication skills;
6. Demonstrated proficiency in written communication skills;
7. and, Demonstrated ability to work with diverse groups/populations.

Sample Position: Data Scientist

Summary of Position Responsibilities: Data Transfer / Data Cleaning / Data Import, Matching/Pulling Data, Process Development, Data Security / Server Administration, and Relationship Building.

Primary Purpose of Position: The Data Scientist will be the lead data scientist for institution. His or her accountabilities will include a broad spectrum of responsibilities ranging from low level algorithm development to stakeholder interaction and relation building. The employee will carry out research involving institution data projects, with an emphasis on optimizing the a Community Database, an integrated data system containing social and human data gathered from several public and nonprofit organizations in the region.

Job Responsibilities:

Data Transfer / Data Cleaning / Data Import

1. Take lead role in data transfer operations between community partners and database.
    a. Develop and monitor data transfer protocols, ensuring appropriate measures for safeguarding and de-identifying data records.
    b. Manage technical consultant(s).
2. Perform cleaning of complex data that is received in multiple files and formats from community partners and researchers.
    a. Ensure accuracy and integrity of data.
    b. Perform data transformations as necessary.
    c. Use independent judgment on data cleaning operations.
    d. Manage technical consultant(s) and graduate student(s) as required.
3. Create databases and import relevant data. Merge and/or normalize data as needed.
4. Link data using internal ID data linking system.

Matching / Pulling Data

1. Ensure timely and accurate data retrieval for researchers.
2. Develop algorithms for interagency record matching based on multiple criteria.
3. Develop algorithms for de-normalizing data for presentation to stakeholders.

Process Development

1. Design, develop and refine processes for operations including but not limited to:
    a. Conventional data pulls
    b. Codebook development and maintenance
    c. Typical cleaning procedures
2. Stay abreast of current technologies and best practices.
3. Design and/or implement data linking system.
4. Write technical documents and make technical presentations as needed.
5. Develop software using R programming.

Data Security / Server Administration

1. Ensure strict adherence to data governance procedures.
2. Ensure that only non-identifiable data is given to researches.
3. Ensure secure data transfers from partner agencies are carried out
4. Act as liaison to IT Services to guarantee the proper administration of secure database and data transfer servers.

Relationship Building

1. Cultivate and maintain working relationships with:
2. Exercise independent judgment and original thinking in support of research that advances the University's reputation.
3. Contribute to proposals for externally funded projects in support of research mission.
4. Participate in departmental seminars and other professional activities
5. Participate in data oversight committee and data quality review meetings.

Education, Skills and Experience Requirements:

Minimum Education/Experience

1. Graduate degree in Computer Science, Data Science, or Math.

2. Professional skills, work experience, and proficiencies required:

- Expertise in Social Research methodology and analysis
- At least 10+ years professional experience.
- Proven background in software development using high level programming languages such as Java or Python.
- Familiarity and comfort with complex algorithm development. Exposure to matching algorithms such as fuzzy matching, soundex, or Levenshtein Distance is preferred.
- Strong communication skills, including the ability to explain technical concepts to nontechnical audiences, to clearly document work, and to interact with users to ascertain project requirements for translation into technical specifications.
- Background in flowcharting and other system diagramming tools.
- Unified Modelling Language, to facilitate technical communications.
- Ability to work independently within a diverse team environment, to effectively plan and prioritize own work, and to adapt to emergent needs and changes in project timelines, as required.
- Expert knowledge in federated database management systems
- Experience in documenting processes.
- Ability and willingness to continually expand and apply new technical skills and knowledge.
- Familiarity with ethical considerations and state and federal regulations surrounding data management, including FERPA and HIPAA.
- Familiarity with statistical software, including SPSS and/or SAS.
- Experience with programming languages, such as R and/or Python, that are relevant to data science are particularly helpful.
- Experience with database development and architecture and best practices using different schema models as well as relational database concepts such as normalization/denormalization of tables.
- Understanding of data security protocols.
- Deep proficiency in SQL query development and optimizations – including joins, stored procedures, views.
- Knowledge of backend web development programming in PHP and familiarity with client side web technologies such as JavaScript and HTML5 and various extensions (AJAX, JQuery, etc.).

3. Ability to design and implement procedures and processes for complex data manipulation operations.

4. Demonstrated ability to multitask and juggle multiple projects simultaneously.

5. Ability to manage graduate students and technology consultants.

6. Readiness to participate in budget creation.

7. Ability to collaborate with other technology groups, such as UNC Charlotte IT Services, to development new secure platforms for data transfer, database administration, data migration, and backup and recovery procedures.

8. Willingness to act in an advisory capacity to other technologists and information specialists.

Preferred Education, Knowledge, Skills and Experience

- PhD degree in Computer Science, Data Science, or Math preferred.

Sample Position: Database Administrator

Key Role: Database Administrator

Act as the team's technical expert responsible for configuring and managing critical business continuity, mass notification, and situational awareness tools and systems. Develop and apply quality assurance processes to optimize system performance and ensure data accuracy and integrity. Develop, establish, and manage ongoing enhancements to systems and interfaces to enhance effectiveness of the overall architecture. Design and create plans, reports, and dashboards using Fusion and other tools. Establish and maintain user accounts and privileges. Prepare operational and training documents and train and support end users.

Basic Qualifications:

1. 3+ years of experience with application systems administration and designing and managing applications
2. 2+ years of experience with formal IT project management
3. 2+ years of experience in working with systems development life cycle processes
4. Ability to pay strict attention to detail
5. Ability to obtain a security clearance
6. BA or BS degree

Additional Qualifications:

1. Experience with designing and managing distributed databases
2. Experience with designing and creating reports
3. Experience with supporting business continuity or continuity of operations programs, including plan development, maintenance, and testing
4. Experience with SQL
5. Possession of excellent technical problem-solving skills
6. Possession of excellent oral and written communication skills

Sample Position: Data Integration Specialist

Job Description: The Data Integration Specialist works collaboratively with other members of the IT team in setting standards for database structures and integrations that enable advanced reporting and insights. Primary responsibilities will include creating and managing the integration of data collection tools and providing technical support to end users. Workflows and standards created by the Data Integration Specialist will be used widely to move and transform data, preparing it for reporting and further analysis and consumption.

Responsibilities:

1. Establishes standards for workflow design as well as database structures, based upon feedback from colleagues, industry trends, and best practices.
2. Creates and manages new database structures (fields, layouts, views, etc.) and validates collected data using other sources to assist the reporting and analysis of survey data.
3. Administers and configures integration of survey tools.
4. Collaborates with other members of the IT team to ensure that integrations align with existing platform architecture and strategies.
5. Serves as a main IT point of contact for Evaluation team related to data collection and integration.
6. Communicates technical concepts to other staff and key stakeholders who have little to no technical ability.
7. Develops and documents IT operational processes, methodologies and associated training related to system integrations.
8. May conduct software needs assessment(s) and make recommendations to meet business needs and optimize data integration.

Requirements:

1. BS/BA required
2. Experience working within a complex integration environment.
3. Experience with designing data storage solutions for reporting purposes and ensuring data integrity or “cleanliness.”
4. Possesses technical and quantitative reasoning skills for understanding survey questionnaires, research methodologies, and the purpose of the research.
5. Familiarity with survey data variable types and nuances.
6. Experience with CRM and/or relational databases.
7. Experience with data collection tools.
8. Active learner who adopts new tools and work processes quickly and solves problems creatively and independently. Ability to anticipate risks and devise solutions in the moment.
9. Ability to manage one’s time and multiple projects with a positive, flexible attitude in a complex and rapidly changing work environment
10. Outstanding written and oral communication skills with an ability to articulate and synthesize information in a compelling and organized manner.

Actionable Intelligence for Social Policy (AISP)

Housed at the University of Pennsylvania, AISP works with state and local governments to develop Integrated Data Systems (IDS) that link administrative data across government agencies. IDS give governments and their partners the ability to better understand the needs of individuals and communities and improve programs and practices through evidence-based collaboration. For more information, please visit AISP's website.

MetroLab Network

In 2017, MetroLab launched its Data Science and Human Services Lab, as an effort to bring together academics, city and county practitioners, and non-profit leaders to consider the issues at the intersection of technology, analytics solutions, and human services deployment. Last year, we published our first report on the topic: "First, Do No Harm: Ethical Guidelines for Applying Predictive Tools Within Human Services" which addresses the use of predictive risk modeling in human services. For more information, please visit MetroLab Network's website.

National Neighborhood Indicators Partnership

Since 2013, NNIP has been involved with promoting the use of and expanding access to Integrated Data Systems. Through a Casey funded project they had six partners explore the use of IDS to combine info from IDS with place/neighborhood data to address a local policy question. Since the end of the project, they have continued to convene their partners around this issue and provide a forum for peer learning and exchange. For more information, please visit NNIP's website.

CONTACT:
info@metrolabnetwork.org
MetroLab Network
444 N. Capitol Street, NW, Suite 399
Washington, DC 20001

@metrolabnetwork