Data Modernization Recommendations for Water Sector Public Agencies

Duke
NICHOLAS INSTITUTE
FOR ENVIRONMENTAL POLICY SOLUTIONS

Internet of Water

CONTENTS

INTRODUCTION

In 2016, the Water Policy Program (WPP) at the Nicholas Institute for Environmental Policy Solutions at Duke University launched a nationwide effort to modernize our nation's water data infrastructure. The Internet of Water project (IoW) began with a dialogue series in partnership with the Aspen Institute and has evolved into close engagements with those in the water sector at every level, from federal to state to local policy makers, to consultants and engineers, to local water managers in both urban and rural communities, including public water utility operators and managers, to irrigators, farmers, and ranchers. The results of these engagements are now implemented in four key states (California, New Mexico, North Carolina, and Texas) in a variety of pilot programs. A common theme among these pilots is the need for and challenges surrounding data infrastructure modernization. What is clear from this work is that resilient water management requires modern data infrastructure.

The following Implementation Guide is the result of the WPP's Technology Adoption Research Project. The Technology Adoption Research Project follows 18 months (2019-2021) of pilot engagements and focus groups conducted by the WPP IoW team at Duke University. In addition to already existing best practices in the field of digital transformation, the engagements from our four pilot states provided an opportunity to work closely with public agency partners in these states, and to observe how their agencies adopt new technology, and in particular, their challenges in doing so. We also conducted several focus groups to talk directly with a diverse range of public agency employees about their experiences working with data in their agency, challenges and barriers to improving water data infrastructure, their observed benefits and successes around modernization efforts, and recommendations on how they feel their agencies can modernize.

The results of these best practices and observations follow, along with a detailed roadmap of recommended actions that public agencies can take to implement the Water Data Sharing Pilots outlined in the 2021 Infrastructure Bill, and finally, estimated costs associated with such activities, modeled from the WPP's IoW pilot projects.

CONSIDERATIONS FOR PUBLIC AGENCIES

Generally, challenges with technology adoption and modernization in public agencies are rooted in four causes:

1. **Lack of Demand:** In the private sector a demand signal, powered by the market, indicates a desire for modernization. In the public sector, no such demand signal exists to drive action; instead the driver for action is the need to deliver the 'public good' in a manner that is effective and efficient.
2. **Necessary Transparency:** The private sector is not necessarily better at technology development. When the private sector makes a mistake or fails at technology development, it is not public knowledge. However, in the public sector, transparency requires that attempts and failures are public knowledge, often leading to questions surrounding appropriate use of public funds. Therefore, public agencies and the people that lead them are often **risk averse** and incentivized to maintain the status quo.

NOTABLE TERMS

Digital or data transformation: a fundamental change in how organizations think about, collect, and manage data, resulting in the modernization of data into a service rather than a single-use product.

Technology adoption: the implementation of the technological systems necessary to modernize an organization's data systems

Modernization: to bring a process, organizational structure, regulation, or mission up to a current standard. While standards do evolve over time, modernization does not necessarily mean "new." Modernization also does not mean "digital," as there are some processes that are not accessible or improved by digital formats (for example, services for populations without easy access to the Internet). It is also important not to equate modern with "permanent," as truly modern systems are those that resilient to contextual changes. A modern system should constantly reassess how well it responds to the changing context around it. This adaptability makes modern systems simple, usable, useful, reliable, and resilient.

3. **Competing Priorities:** Private organizations can identify and focus on a priority based on a market-driven organizational mission, which can be revised and changed in response to market demand. In contrast, the mission of public agencies is often codified in law and often more expansive than those from private agencies. This can create competing priorities that make cross agency coordination or centralized management difficult.

4. **Generational Conflict:** Public agencies often have multiple "generations of technology," under one roof. This means that cultural conflicts are the cause of much resistance to technology adoption and modernization within and between agencies. Today's systems are not only built upon legacy technologies, but also the *thinking* that created them.

Because of the challenges specific to public agencies, it is critical to the success of modernization efforts that technical frameworks and implementation plans developed through the process are done so within the context, capacities, and capabilities of the public agency. Modernization efforts require coordination across the divisions of the public agency and should be carried out in accordance with the organizational structures that commonly exist.

RECOMMENDATIONS

Best practices in the field of digital transformation, lessons learned from community and public agency engagement during the pilot period of the IoW project, as well as survey and interviews conducted during the Technology Adoption Research Project, provide the basis for the following recommendations:

1. **Identify and provide incentives for data modernization.** Because human capacity, digital infrastructure, and financial resources are limited for public agencies, an important mechanism for water data infrastructure modernization is to tie grant or other funds to the development of and compliance with standards for improved accessibility, interoperability, and modernization of public agency water data infrastructure. This would include grant programs offered by federal agencies, but also should include grants provided through philanthropic and other non-profit organizations. Documentation of standards and best practices for data modernization should be provided to grantors as guidance for incorporation into award requirements. While cultural barriers will not be resolved with incentives alone, the application of funds toward modernization can provide in-house demonstrations that can be persuasive to resistant leadership, particularly when

they result in improved employee efficiency. These demonstrations can also be used by agency leadership to champion water data modernization and urge policymakers to develop sustainable funding sources for these initiatives.

2. **Connect information delivery with policy outcomes.** While leaders often tout "data-driven decision-making," participants in the study and pilot period engagements struggled to provide evidence of decisions directly informed by data. This may be the result of cultural or behavioral process in which decision-makers, who have historically not had access to information delivered in meaningful ways, continue to fall back on their traditional methods of decision-making. To understand how data are informing decision-making, agencies need to ensure clear avenues of information delivery are accessible to decision and policymakers, then seek out and compile evidence about how their data modernization efforts have directly improved or informed decisions. These findings will promote further support of data modernization, as those who are tasked with resource allocation are also directly benefiting from the modernization process.

3. **Resolve issues with procurement processes.** Guidelines should be developed for agency procurement contracts to ensure that contractors follow modern data principles, meet agency needs for sharing and interoperability, and build in plans for sustained maintenance and maturity of systems.

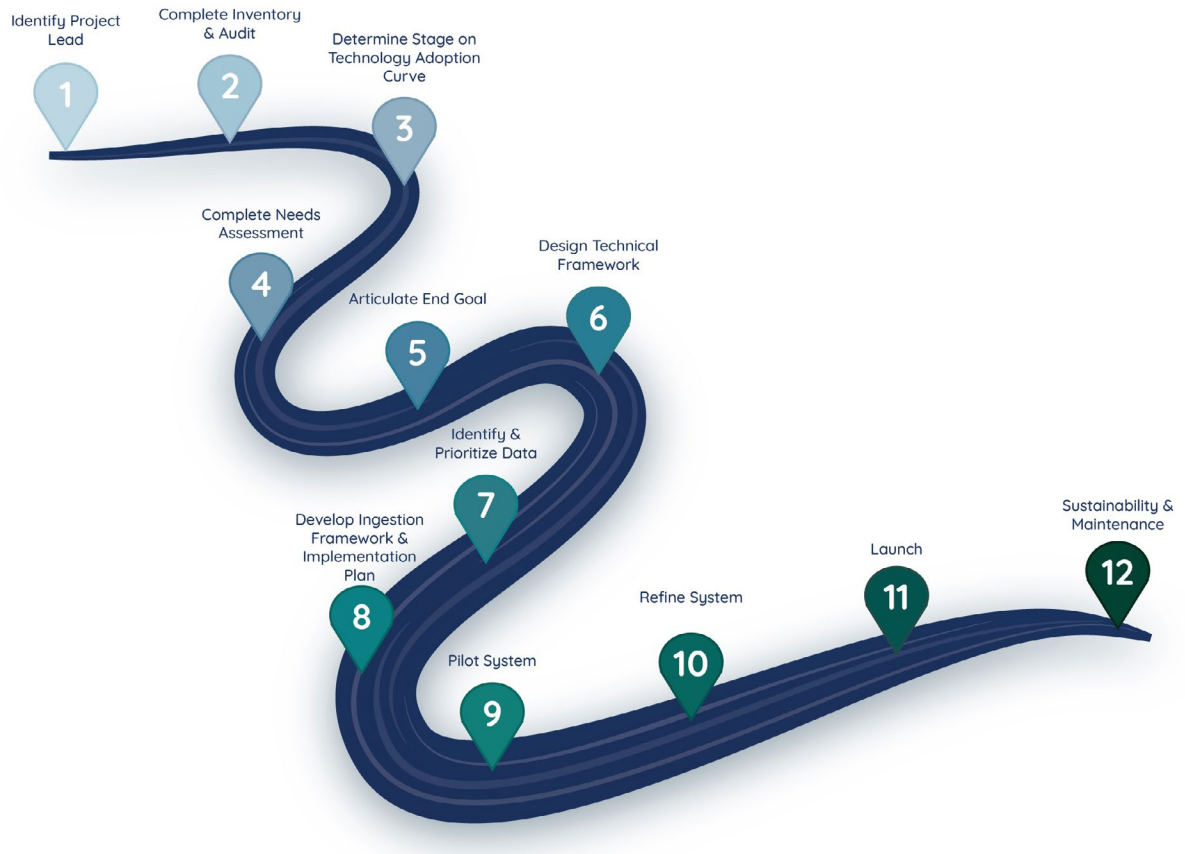
4. **Identify leader for cross agency compliance and establishment of standards.** Executed contracts should be overseen by a Project Lead, as described in Observations and Lessons Learned #5 of the Technology Adoption Research Project Report.

5. **Invest in modernization and technology adoption training for water leaders.** Communication and training programs designed to inform decision and policymakers about how to interpret and understand data, apply data to decision-making, and appreciate the need for and power of modern data infrastructure will narrow the cultural divide between different generations of technology, equip leaders with the knowledge they need to engage with their agency staff about modernized data infrastructure, and remove much of the resistance and fear over technology adoption. Modeled after the Harvard Evidence for Policy Design program, water data modernization and technology adoption training for water leaders would deploy teams to states for in-situ training. Funding for such a program would come from a combination of public grant and philanthropic funds.

ROADMAP FOR MODERNIZING WATER DATA INFRASTRUCTURE

The following roadmap incorporates best practices for data ingestion, adapted from private industry (See Snowflake, Striim, and Qlik), and incorporates agile development guidance from the U.S. Digital Services Playbook. Additionally, the following represents public agency-specific guidance developed during the IoW's pilot studies and the observations and lessons learned from the Technology Adoption Research Project, all in accordance with the IoW Principles.

Technology Adoption Roadmap



1. **Identify Project Lead**
    - Assign a project lead(s) responsible for the implementation of the water data modernization effort. This person(s) should have technical project management experience and appropriate knowledge of water data to

navigate engagement with users and stakeholders.

2. **Complete Asset Mapping, Inventory, and Audit**
    - Conduct asset mapping to understand the different skills, capacities, and capabilities from different teams across agencies or agency divisions. Understanding in-house capacity is important in creating an efficient, effective, and sustainable modern data infrastructure.
    - Conduct an inventory of current systems and platforms. (See the IoW's Water Data Inventory guidance).
    - Fundamental questions during this process are what does the existing system look like? What technical skills currently exist within the agency? What capacity do those with technical skills have to devote to modernization efforts?

3. **Determine Stage on Technology Adoption Curve**
    - In reference to the Technology Adoption Framework (Appendix A) determine current agency location along the technology adoption curve.

4. **Complete Needs Assessment**
    - Conduct internal engagement regarding barriers or challenges to movement along the technology adoption curve and identify internal and external resources that could be allocated to the modernization effort.

5. **Articulate End Stage Goal**
    - Set an end-goal for the stage on the Technology Adoption Curve. This determination should be based upon the starting point, needs assessment, and agency capacity and capability as identified in previous steps of the roadmap. Not every agency will move directly to Stage 4, for example. It is important to identify a realistic goal and end stage for data modernization.

6. **Design Technical Framework**
    - A technical framework document defines specifications and implementation for data modernization. The IoW recommends the following specifications:
        - » Metadata is published on the web, ideally in compliance with data-on-the-web best practices from W3C
        - » Data is available for download in bulk and/or API in OPEN, non-proprietary formats
        - » To the extent possible, data bulk download formats and/or APIs will follow community standard patterns (e.g. OGC standards); metadata

will be included with data and of sufficient quality for users to make judgments as to what purposes the data is fit for use; and data content will reference publicly available definitions, controlled vocabularies, and data standards appropriate to the data's subject matter

    - » Data will be published and identified with version records and made available (to authorized users) so that workflows can be reproduced
    - » Open formats data content standards, and data exchange or API standards for similar kinds of data should reference community, national, or international standards where practicable (See IoW Data 101 Guidebook).

- A technical framework should also:
    - » Be informed by engagement and needs assessment (Step 4)
    - » Responsive to current location on technology adoption curve and desired end stage (Step 5)
    - » Outline data standards, metadata standards, and software needs and acquisition plan (Step 6a)

7. **Identify and Prioritize Data**

    - Many public agencies hold large amounts of data, covering decades in time. Successfully modernizing data infrastructure requires a strategic approach to identify and prioritize data to be incorporated into a newly modernized data infrastructure. This is particularly important for legacy data. Not all legacy data need be digitized; therefore, datasets of most need should be prioritized for digitization. In addition, datasets that are commonly shared internally or externally should be prioritized for incorporation into a newly modernized infrastructure to address issues of version control and challenges with cross-agency collaboration as outlined in the Technology Adoption Project Report.

8. **Develop an Ingestion Framework and Implementation Plan**

    - An ingestion framework is a process for transporting data from various sources to a storage repository or data processing tool (See Snowflake, Striim, and Qlik). Data ingestion processes should be developed based on the data architecture, the volume of data to be ingested, and the frequency of data ingestion. An Ingestion Framework articulates these processes as well as any integration challenges (such as data compatibility and standardization) that are required for successful data modernization.
    - An Implementation Plan is a document that articulates an organizational

strategy for the execution and sustainability of a data modernization effort. An implementation plan consists of:

- an engagement strategy (how to identify and engage with targeted stakeholders or users of the newly modernized data system – these may be internal or external stakeholders),
- long term care and maintenance plan for the resulting systems or products,
- privacy restrictions and guidelines,
- funding requirements,
- associated staffing needs, and
- monitoring and evaluation strategy for impact assessment.

9. **Pilot System**
    - In accordance with agile development best practices (See U.S. Digital Services Playbook), the development of new systems must include engagement with users and stakeholders to assess usability, functionality, and efficiency of the modernized system.
    - Steps 9 and 10 articulate an iterative process for engagement and refinement. These steps should be repeated until such point in which the new system or product meets user expectations.

10. **Refine System**
    - Based on feedback and lessons learned from the engagement in Step 9, refine the system.
    - Return to Step 9 to further refine.

11. **Launch**
    - Promote the system through internal and/or external communications and trainings to ensure that it is widely adopted by agency staff.

12. **Sustain, Maintain, and Improve**
    - Perform routine maintenance to ensure system is sustained over time.
    - At intervals defined in the implementation strategy, evaluate the newly modernized system, measuring and articulating impact and identifying opportunities for improvement.

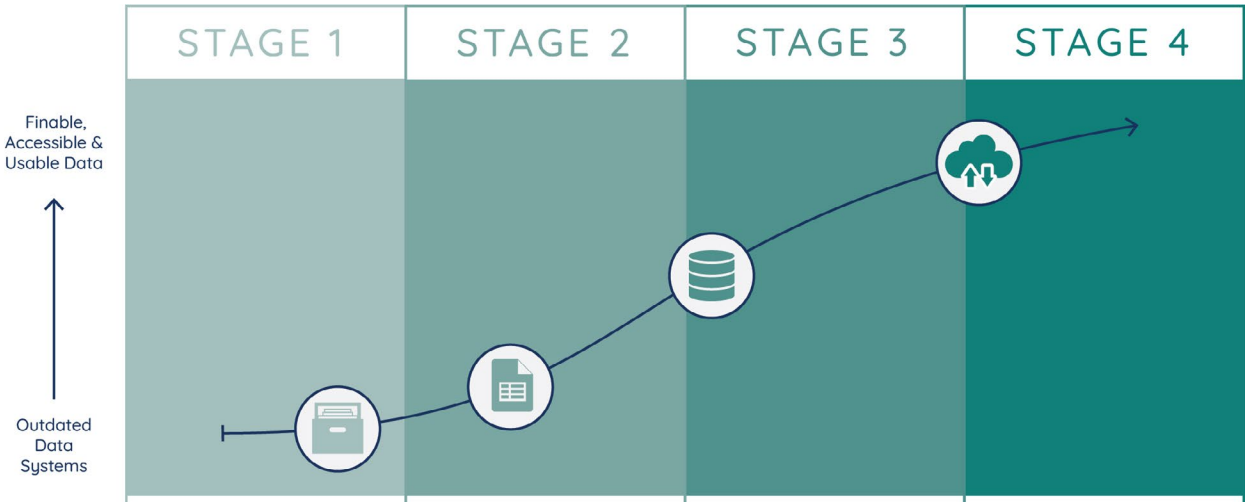
BUDGETING FOR WATER DATA MODERNIZATION

In 2019, the New Mexico state legislature passed the New Mexico Water Data Act. The act established the Water Data Initiative within the New Mexico Bureau of Geology and Mineral Resources. A steering committee of representatives from partner agencies decided on a tripartite implementation structure, with specific responsibilities assigned to various work groups. User engagement was a critical component of the Water Data Initiative, designed to inform understanding of New Mexico agency capacity, assess needs of both agencies and stakeholders, and guide the development of public-facing data platforms. The New Mexico Water Data Initiative (NMWDI) provides a reliable model for other, similar initiatives because it required cross-collaboration between state agencies, funded a lead agency, incorporated stakeholder engagement, and resulted in a public facing platform to facilitate data integration and interoperability.

The initial funding provided for the NMWDI was $435,000 per year for years 1 and 2 of the initiative. The estimated cost for years 3-5 is an additional $500,000 annually to fully support an IT and operations team with up to four full-time dedicated staff who will develop and maintain the cyberinfrastructure and connections to data producers and users. The six other agencies named in the Water Data Act also requested funding. Their funding requests varied significantly depending upon existing agency capacity, existing data infrastructure, and the volume of data managed by the agency. Initial investments to modernize data infrastructure for these agencies averaged $410,000, and recurring annual costs averaged $421,330.

Using the New Mexico Water Data Initiative as a guideline, an estimated cost for implementation is $450,000 annual startup costs for the lead agency and $410,000 annual startup costs for each participating agency. Recurring costs are an estimated $500,000 per year for 5 years for the lead agency and $430,000 per year for 5 years for each participating agency.

APPENDIX A: TECHNOLOGY ADOPTION CURVE



STAGE 1	STAGE 2	STAGE 3	STAGE 4
• Most data are collected in paper formats and stored in filing cabinets. • Staff manually access and compile data. • Staff can not create data visualizations or analytics • Data are not standarized • Data quality is not documented • Staff cannot integrate their data with data from other divisions or agencies • It is difficult to share data across or between agencies.	• Most data are collected in digital formats and stored on local computers. • Staff manually access and integrate data. • It is difficult for staff to create data visualizations and analytics • Data are not standarized • Data quality is documented in an unstandarized way that is not legible to those outside the division or agency • It is difficult for staff to integrate their data with data from other divisions or agencies • It is difficult to share data across or between agencies.	• Most data are collected in digital formats and stored in a shared database/s. • Staff access data through one or more databases and manually integrate data. • It is difficult for staff to create data visualizations and analytics • Data may not be standarized across divisions or agencies • Data quality is well documented • It is difficult for staff to integrate their data with data from other divisions or agencies • It may be difficult to share data across or between agencies.	• Data are collected in digital, machine-readable formats and stored in a shared database. • Staff access and integrate data through APIs. • Staff can easily create data visualizations and analytics • Data are standarized across divisions and agencies • Data quality is high and is well documented • Staff can easily integrate their data with data from other divisions or agencies • It is easy to share data across or between agencies.

APPENDIX B: IOW PRINCIPLES

The following principles were originally developed during the Aspen Institute Dialog Series on Water Data, and published in the 2017 report "The Internet of Water: Sharing and Integrating Water Data for Sustainability." In 2021, the principles were revised in consultation with the Internet of Water Advisory Board to reflect lessons learned over the first three years of project implementation.

1. Water data are essential for efficient, equitable, sustainable, and resilient water planning, management, and stewardship.

2. Modern data infrastructure increases the usefulness of water data and enables its broadest possible application.

3. Data equity is necessary for water equity; modern data infrastructure should be implemented and governed so that data are usable by and for overburdened communities.

4. All water data produced for the public good should, by default, be findable, accessible, interoperable, and reusable (FAIR) for public use or authorized users.[1]

5. Security and privacy risks associated with sharing data can be mitigated using mechanisms for tiered access for authorized users.

6. Commonly accepted data, metadata, and exchange standards should be adopted by water data producers to promote interoperability, efficiency, sharing, equity, and secondary uses of data.

7. Control and responsibility over data are best maintained by data producers.

8. Data producers are responsible for sharing data of known quality and documenting essential metadata; data users are responsible for determining whether data are appropriate for specific purposes and uses.

9. Federated, distributed systems of interoperable public water data generally provide scalability and flexibility to meet the diverse needs of data producers and users.

DEFINITIONS

Accessible: Full data sets are available to the public or authorized users for download in machine-readable, non-proprietary formats.

Authorized users: The group of users that are allowed to access a given dataset. The default group of authorized users for public water data is the general public. In certain cases, such as datasets that include personally identifiable information or that represent serious security risks, this group may be limited by data producers to users with specific data use agreements or security clearances.

Data hubs: Structured sources of standardized water data aggregated by theme or geography.

Data producers: Entities that collect data for a specific purpose and have authority over what and how data are produced, including organizations that manage citizen science and crowd-sourced data (e.g., a wastewater treatment plant that produces data about surface water conditions, a state agency that holds water rights data, a non-governmental organization (NGO) that collects water data samples, a private company that takes meter readings).

Data standards: Guidelines regarding how data about a particular topic is (1) structured, defining what data elements should be present; (2) populated, defining the kind and quality of information represented; (3) encoded in machine-readable formats; and (4) made interoperable for data exchange.

Data users: Primary and secondary entities that use water data to create information and value. Primary users are the producers who use the data they collect to meet a specific mission (e.g., a state environmental quality agency that regulates discharges of pollutants, a reservoir operator that regulates the flow of water through a dam). Secondary users create value by combining multiple types of data, typically from multiple organizations (e.g., a conservation organization building stream restoration maps from data held by a utility, state, and reservoir operator; a private company assessing, modeling, and visualizing the environmental impacts of real estate development).

Findable: Data and metadata published on the web in compliance with data-on-the-web best practices, ideally tied to a common hydrography.

Interoperable: Data bulk download formats and application programming interfaces (APIs) that follow community standard patterns; metadata are included with data and of sufficient quality for users to make judgments as to what purposes the data is fit for use; and data content references including publicly available definitions, controlled vocabularies, and data standards appropriate to the data's subject matter.

Metadata: Metadata is information about data that assists potential data users in the discovery, access, and use of the data. It can describe the identity, subject matter, and producer of the data to aid in data discovery. It can describe the location, license, and point of contact for the data to assist in data administration and access. It can describe the structure, format, and any applicable data standards to assist in the use and manipulation of the data.

Modern data infrastructure: An integrated system of 21st-century information technologies, which includes common standards, formats, and tools designed to make water data easy to find, access, and share online. This system is connected by a network of people and organizations serving as water data producers, users, and hubs.

Overburdened community: Minority, low-income, tribal, or indigenous populations or geographic locations in the United States that potentially experience disproportionate environmental harms and risks. This disproportionality can be as a result of greater vulnerability to environmental hazards, lack of opportunity for public participation, or other factors. Increased vulnerability may be attributable to an accumulation of negative or lack of positive environmental, health, economic, or social conditions within these populations or places. The term describes situations where multiple factors, including both environmental and socio-economic stressors, may act cumulatively to affect health and the environment and contribute to persistent environmental health disparities.[2]

Water data produced for the public good: refers to water data collected for any public mission or purpose, including for regulatory compliance, either made available to the public or limited to authorized users.

Reusable: Data that is published and identified with version records and made available to the public or authorized users so that workflows can be reproduced.

1 Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

2 US Environmental Protection Agency. EJ 2020 Glossary. https://www.epa.gov/environmentaljustice/ej-2020-glossary

RECOMMENDED READING

Bellotti, Marianne. *Kill it with Fire: Manage Aging Computer Systems (and Future Proof Modern Ones)*. No Starch Press: San Francisco. 2021.

Carnahan, R., Hart, R., Jaquith, W. "State Software Budgeting Handbook." 18F Technology Transformation Service, General Services Administration: https://derisking-guide.18f.gov/state-field-guide/

Fenton, N., Pandel, A., Harrell, C. "Best Practices in Government Digital Transformation." https://github.com/18F/transformation-research

Ferguson, C. "What went wrong with America's $44 million vaccine data system?" MIT Technology Review, January 30, 2021: https://www.technologyreview.com/2021/01/30/1017086/cdc-44-million-vaccine-data-vams-problems/

Ferguson, C. "Why is it so hard to build government technology?" MIT Technology Review, March 17, 2021: https://www.technologyreview.com/2021/03/17/1020811/better-tech-government-pandemic-united-states/

Harrell, Cyd. *A Civic Technologist's Practice Guide*. Five Seven Five Books: San Francisco. 2020.

Karman, P. "Digital Service is not about technology." February 15, 2020: https://peknet.com/2020/02/15/digital-service-is-not-about-technology/

Ogle, A. and Eaves, D. "Moved by Meaning: Attracting, retaining, and managing digital talent in government." Open Leadership Camp Working Paper Series: file:///Users/aw75/Downloads/SSRN-id3089936.pdf

Pahlka, J. "Delivery-Driven Government: Principles and Practices for Government in the Digital Age." Code for America. https://medium.com/code-for-america/delivery-driven-government-67e698c57c7b

Pahlka, J. "Delivery-Driven Policy: Policy Designed for the Digital Age." Code for America: https://www.codeforamerica.org/news/delivery-driven-policy/

Shevin, M. "How public-interest technologists can build government that works for everyone." Fast Company 02/12/2021: https://www.fastcompany.com/90604086/public-interest-technology-government

A special thanks to the Executive Education program, "Digital Fundamentals for Public Agencies" by Georgetown University's Beeck Center for Social Impact and Innovation and the McCourt School of Public Policy.