STATE OF NEW HAMPSHIRE
BEFORE THE
PUBLIC UTILITIES COMMISSION

In the matter of

Electric and Natural Gas Utilities

Docket No. DE 19-197

Development of a Statewide, Multi-Use Online Energy Data Platform

PREFILED DIRECT TESTIMONY

OF

JAMES BRENNAN
FINANCE DIRECTOR

August 17, 2020

Table of Contents

I. INTRODUCTION

A. Background

Q. Please state your name, business address and current position.

A. My name is James Brennan. I am the Finance Director at the Office of the Consumer Advocate (OCA). My business address is 21 South Fruit Street, Suite 18, Concord, New Hampshire.

Q. Please describe your education and professional experience.

A. I earned a Bachelor degree from Saint Bonaventure University and an MBA in Finance at Syracuse University and completed a nine month JP Morgan Chase (formerly Chemical Bank) MBA Management Training Program. I have completed additional courses in business, finance, software development, electric utility regulation, regulatory finance and accounting, and Smart Grid.

In my present position at the OCA I perform economic and financial analysis of utility filings across all industries, draft discovery and testimony, and provide guidance on financial policy and regulatory issues.

My business career began in banking as First Vice President at Chemical Bank, 1980-1989, with responsibilities as analyst, credit department manager, account relationship manager, and course designer and instructor of Risk Assessment training. I have experience managing business and technology operations. At TD Waterhouse Securities, 1995-2001, I had direct management and budget responsibility for vendor management, application development, New York Stock Exchange Compliance, and month-end data processing of the third largest Wall Street brokerage statement operation during a four-year period of 400 percent growth and a

Number 1 ranking by Smart Money.

I have experience in IT project management and software design. I have designed systems at Mathematica Policy Research in Princeton, New Jersey, a national data collection system for the U.S. Department of Labor that analyzed training program results and generated performance metrics used by the Office of Management and Budget. I oversaw deployment of systems to 700+ locations nationally and to a centralized platform in Washington D.C. at the U.S. Department of Labor datacenter. I have implemented paperless technology systems for Waterhouse Security National Investor Clearing Corporation stock clearing operation (2000), managed the launch of an eServices web site providing on-line secure for Waterhouse Securities, designed Microsoft.NET and SQL Server based software systems for Mathematica Policy Research, and project-led business and development teams deploying cloud based Microsoft Customer Relationship Management solutions for Southern New Hampshire University. I have designed and taught courses in Corporate Finance, Microsoft applications and Microsoft C# programming language.

My experience in the utility energy sector includes the role of Smart Grid Analyst for the New Hampshire Public Utilities Commission (PUC), involvement as a member of the Smart Grid Interoperability Panel-Cyber Security Working Group (SGIP CSWG) which contributed to the NIST Interagency Report (NISTIR) 7628, and worked on the Advanced Meter Infrastructure (AMI) CSWG. I have testified on technology, advanced meter infrastructures, consumer enabling data systems, and energy utility dockets before the PUC and the New Hampshire Legislature. I serve on the board of the Smart Energy Consumer Collaborative (SECC) and as a member of its education and research committees. The mission of the SECC is to understand the needs and wants of electric utility customers.

Q. Have you previously provided testimony before the New Hampshire Public Utilities Commission?

A. Yes.

Q. In which dockets did you testify?

A. I provided testimony before the Commission in the following dockets:

- DE 10-055 Unitil, Inc., rate case testimony assessing the company's smart grid investments;
- DE 13-177 Public Service Company of New Hampshire (PSNH), testimony regarding Least Cost Integrated Resource Planning;
- DE 14-120 Public Service Company of New Hampshire (PSNH), testimony on reconciliation of the company's energy service costs;
- DW 13-130 Pennichuck Water Works, Inc., rate case addressing the company's revenue deficiency;
- DG 15-090 Northern Utilities, Inc., testimony on design of interstate pipeline refund in cost of gas rates;
- DE 11-250 Public Service Company of New Hampshire (PSNH), testimony (adopted) on investigation of Merrimack Station scrubber project cost recovery;
- DE 14-238 Public Service Company of New Hampshire (PSNH), testimony on divestiture of PSNH generation assets;
- DE 15-137 Energy Efficiency Resource Standard, testimony on utilities empowering residential customer through modern electronic data platforms;
- DE 16-384 Unitil Energy Systems Inc., testimony on company pilot to design a utility energy data sharing platform;
- DG 16-383 Liberty Utilities Granit State Electric; testimony regarding long term trend and benchmark analysis using FERC data.
- DE 15-464 Approval of Lease Agreement between Eversource and Northern Pass Transmission LLC, testimony on land and lease valuation

Q. Have you provided public comments to the Commission?

A. Yes, I provided public comments in IR 15-296, the Commission's ongoing Grid Modernization investigation, in which I discussed the definition and elements of grid modernization.

B. DE 16-384 Data Working Group

Q. What was the DE 16-384 Data Working Group and why are you discussing it?

A. The DE 16-384 Data Working Group (DWG) arose as a result of the DE 16-384 Unitil Rate Case Settlement Agreement[1] and functioned during the period of November 2017 to March 2019. The Settlement Agreement called for the creation of an "Energy Data Plan which will include documentation of tables and relationships (Logical Data Model) within the Energy Database."[2] The DWG was comprised of five technical people[3] and included one member from the OCA, two member from PUC Staff, and two members from Unitil. I am discussing the DWG because there was a tremendous amount of work done on the concept of an energy data model that should be leveraged in this docket and that I used to generate work product[4] that greatly informs my testimony.

Q. What was the primary goal of the DE 16-384 DWG?

A. The DWG's primary activity was to design an initial version of a Logical Data Model as stipulated in section 7.7 of the settlement agreement. We believed this was step one in meeting the goal of creating an Energy Data Plan also stipulated in section 7.7. For the development of

[1]Settlement Agreement on Permanent Distribution Rates in Docket No. DE 16-384, Feb. 22, 2017 (Tab 44), available at https://www.puc.nh.gov/Regulatory/Docketbk/2016/16-384/LETTERS-MEMOS-TARIFFS/16-384_2017-02-22_STAFF_SETTLEMENT_AGREEMENT.PDF.

[2]Section 7.7 of the Settlement Agreement, at 14. The text of section 7.7 appears in Appendix 1 at 10.

[3]Collectively, the DWG members possessed senior levels of expertise and experience in the following areas: data architecture and modeling, software design and development, databases (design and administration), utility back office system and architecture, utility operations, engineering (power flow, metering, grid assets and infrastructure), utility analysis, and IT systems analysis, and project management in IT and Operations. A list of members can be found on page 10 of Appendix 1.

[4]The technical documents include: (1) a beta version of a Logical Data Model document, (2) a system design architecture, and (3) a 3-phase project plan which I refer to as "2019 DWG Road Map" that could lead to statewide data sharing for New Hampshire over a three year period.

the data model, the DWG agreed to the philosophy and overarching strategy that the Logical Data Model goal would strive to be future-proof (flexible), generic (multi-utility and multi-fuel/service), standards-based, and multi-use based on six core use cases. With that, the DWG embarked on a highly collaborative analysis of Energy Data to develop a Logical Data Model.

Q. What kind of activities did the 16-384 DWG undertake to achieve its goals, including the creation of a Logical Data Model?

A. The DWG conducted 19 hands-on technical working meetings. Some sessions included additional in-house experts from Unitil (covering focused technical areas – advanced metering infrastructure, utility data systems, cybersecurity, etc.). Some meetings included national experts including ones from the U.S. Department of Energy as well as Chris Villarreal of Plugged In Strategies for guided discussions on issues such as data privacy and data access and experiences in California. Other meetings included Texas utilities with expertise in Smart Meter Texas.

Q. What were some specific topics covered in 19 formal meetings of the DWG?

A. Meeting agendas focused on areas included: (1) analysis of Unitil IT systems at a high-level, (2) analysis of Data Entities including attributes and relationships and interactions, (3) analysis of the ESPI and Green Button data standards, (4) analysis of TOU data architecture, (5) analysis of complex metering scenarios such as net metering, storage, and multi-tenant distributed generation, (6) data privacy and security, (7) data modeling, and (8) manual User Acceptance Testing (UAT) of the Logical Data Model.

C. Introduction of the SB284 Legislation for a Statewide Energy Data Model

Q. What prompted the OCA's effort to pursue legislation for a Statewide Multi-Use Online Energy Data Platform?

A. In the fall of 2018 while the DWG was still engaged, the OCA decided that the substantial progress being made in DWG could provide an effective foundation for a new effort that would elevate the data sharing dialog to a statewide level that included all utilities. The OCA provided an initial draft of the bill and Senator Fuller Clark introduced the measure for consideration in 2019. The Senate adopted the bill, whereupon it moved to the House and was heard by the Committee on Science, Technology, and Energy (STE).

Q. What materials were provided to the Science Technology and Energy Committee to support the idea that a statewide data platform was feasible?

A. The presentation provided to STE includes a synopsis of materials including a copy of a brief on the issues and proposed solution, a summary of the DWG work, and a three phase project plan. I have provided the "STE Presentation and Handout" as Appendix 1.

Q. Please briefly discuss the 3-Phase, 2019 DWG Road Map that influenced the STE prior to passage of SB284.

A. Please see page 7 of Appendix 1 for the high-level but comprehensive initial Energy Data plan that takes New Hampshire from where it was in the spring of 2019 (and where it remains today with major Data Integration Distance), to full statewide data sharing in 2022 which I describe in my testimony as Standardized Integrated Data Sharing.

D. Purpose and Introduction to Key Terms

Q. What is the purpose of your testimony?

A. The purpose of my testimony is to provide a comprehensive set of data sharing recommendations for New Hampshire that meet the intent of Senate Bill 284 as codified at RSA 378:50-54. The goal is to create a data platform that will provide opportunities for long term energy innovations, expanded markets, as well as diversified products and services that the Commission and others have envisioned for a modern grid. My recommendations cover technical, operational and regulatory areas and are both strategic and systemic in nature. The recommendations propose a phased multi-year project planning approach that will reduce risks and cost, while synergistically supporting on-going development of data needs as they are identified, analyzed, and agreed to by stakeholders in future Commission proceedings on grid modernization, energy efficiency, electric vehicles (EV), distributed energy resources (DER), active demand reduction (ADR) and more.

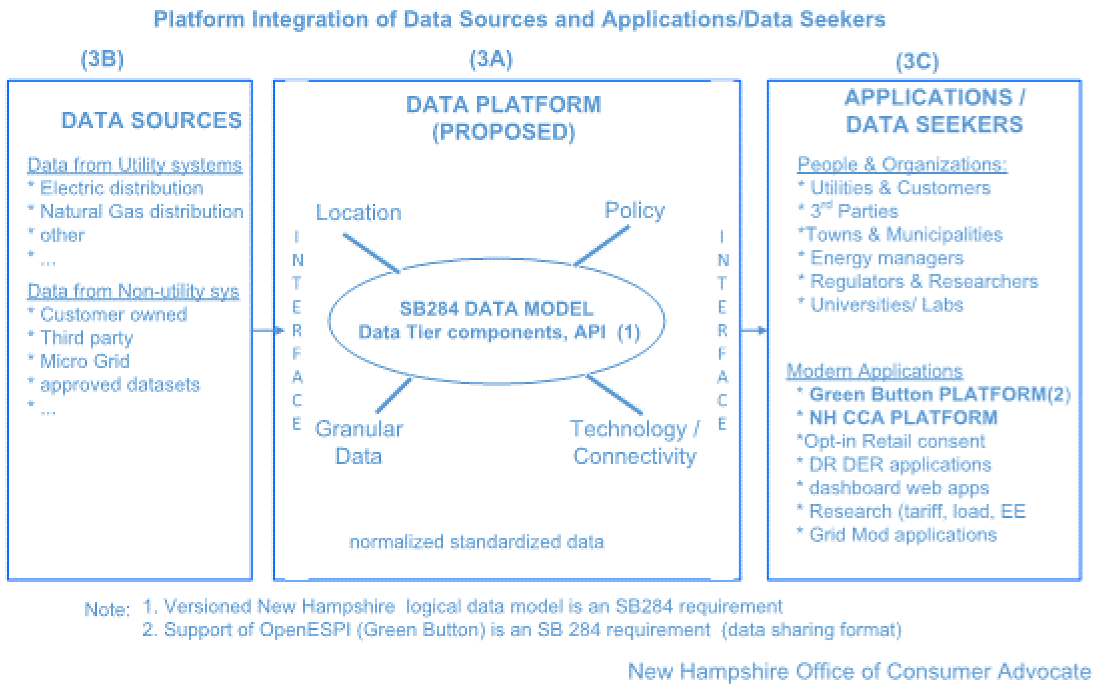
Q. How is your testimony organized?

A. My testimony is organized as follows: The first is an introduction which includes background, the purpose of the testimony and an introduction to key terms. The second section provides a brief history of data standards, evolving data needs, the current issues surrounding data sharing in New Hampshire and my associated vision. Then I describe how the data platform required by the statutory framework should be built so that it solves both the issues of today and can be expanded to meet the needs of the future. Following that, I share views on governance and other issues in front of the Commission.

Q. Are you providing diagrams helpful to the reader to refer to throughout your testimony?

A. Yes. I include a number of figures throughout my testimony. The first is Figure 1, shown below, titled "Platform Integration of Data Sources and Applications/Data Seekers." I will refer to this diagram frequently in my testimony.

Figure 1: Platform Integration of Data Sources and Applications/Data Seekers



A full list of Figures included in my testimony is as follows:

Figure 1: Platform Integration of Data Sources and Applications/Data Seekers
Figure 2: Datasets Represent Data Uploaded to SB284 and Data Shared (Exported) by SB284 via Centralized SB284API
Figure 3: 2020 Forward Looking Data Sharing
Figure 4: Data Integration Distance – Concept
Figure 5: Data Integration Distance – External & Internal
Figure 6: External Integration Strategy - Reduced Data Integration Distance & Statewide Friction
Figure 7: Full Data Sharing Vision – Collaboration of Use Cases in Different Boundaries
Figure 8: Energy Applications & Platforms with Databases – utility and non-utility
Figure 9: Four Categories of Data in NEEDS – from a logical perspective
Figure 10: End-to-End: Zero/Low data Integration Distance
Figure 11: Proposed Build (Including Design Pilot Scope)

Q. Are there any terms you wish to introduce, define or clarify the meaning of relative to your testimony?

A. Yes. Here is a list of terms with descriptions of how I use them in my testimony. These are not full definitions. I encourage both the Commission and others interested in these subjects to seek additional insight by visiting an online glossary of terms from a respected publication.[5]

Application: An Application, also referred to as a "Software Component", is an IT software program running on a computer. An Application or Software Component can do three things: receive data, do something with the data, and send data. Applications allow tasks to be automated by a computer. Applications are designed following some version of a Software Development Life Cycle (SDLC).[6] They are designed and built to satisfy the business requirements developed and approved by owners and stakeholders. Applications are created using programming languages. In this instance they would be object-oriented, which is to say they create references to objects such as meters, premises, or people. The architecture of Applications is service-oriented, which is to say it is structured so that messages request a service such as providing or receiving Energy Data.

Application Programming Interface (API): API and "API Endpoints" are descriptions of software components in terms of operations, inputs and outputs, and underlying types. An API is a set of functions and procedures allowing the creation of External IT Applications that access

[5] National Institute of Science and Technology provides a glossary of terms: https://csrc.nist.gov/glossary.

[6] SDLC https://en.wikipedia.org/wiki/Systems_development_life_cycle.

the features or data of the Data Platform. The proposed SB284 Data Platform will have an API ("SB284 API") that is comprised of a number of API Endpoints that connect to outside Data Seekers and Data Sources. For example, if an External Application requests aggregated data, or if a PUC dashboard asks for all customers on a time-of-use tariff, both External Applications will go to the SB284 API of the Data Platform but they will go to different API Endpoints. Those API Endpoints control the access according to the Data Access Framework policy I am about to define, the Data Privacy Framework policy I also define, and other required standards. An API's functionality can expand over time, for example in order to implement new Use Cases, by developing new API Endpoints. In the OCA Sequence Diagrams (which I am also providing) the SB284 API is an actor represented in all the OCA Sequence Diagrams as the second box from the right at the top of the diagram.

Centralized / Centralized Data: Centralized means bringing control of an activity or organization under a single authority together in one place rather than being siloed. Centralized Data means placing data from different independent data sources (such as individual regulated utilities, competitive energy suppliers, and third parties), as indicated in the Data Sources box on the left hand side of Figure 1, under one umbrella controlled by one Common Data Model, accessed in one place or API, as indicated in the middle box of Figure 1.

Common Data Model: Common Data Model is a Data Model (which I define later in this list) that multiple stakeholders agree to use and follow. Typically it is a published and versioned model. For purposes of my testimony, Common Information Model (CIM) and Energy Service Provider Interface (ESPI) are examples of Common Data Models. They are a means of unifying data meaning and understanding across multiple parties and their IT systems. A "common statewide logical data model" as required by RSA 378:50 is a Common Data Model. Please consider my definition of NEEDS for another example.

"Common Information Model" (CIM): A Common Information Model is a standard for defining device and application characteristics so that system administrators and management programs will be able to control devices and applications from different manufacturers or sources in the same way. By using a CIM, software can be written once and work with many implementations of the common model without complex and costly conversion operations or loss of information. There is an electric utility CIM, developed largely under the aegis of the Electric Power Research Institute (EPRI).[7]

Data Access Framework (DAF) and Data Privacy Framework (DPF): DAF and DPF are centralized policy documents approved by the Commission. The DAF governs who can see what data and under what condition and at what level of granularity and transparency. The DPF governs all privacy rules, consent, opt in, cyber security policy and other protection of customer data. The DAF and DPF represent strictly enforceable rules that limit and guide the design, development, testing, deployment and day-to-day operations of the proposed Data Platform.

Data Entity: A Data Entity can represent an object that exists in the real-world and can be easily distinguished among all other objects of the real world. For example, people can be represented as a Data Entity, as a record in a table in a database. Data Entities can have

[7] *See, e.g.*, the EPRI Common Information Model Primer, available at www.epri.com/research/products/000000003002006001

attributes that define the characteristics of, or the properties of, the Data Entity on a basis which is easily distinguishable among other entities of the real world.

Data Integration or Integration: Data Integration is the ability or the process whereby the Data from one IT system or Software Component is able to flow into a separate IT system or Software Component freely and seamlessly, as though that data had been created in the receiving system and perfectly matches its Data Model. Integration becomes difficult and manual (and therefore costly) when the data (and Data Model) in the source system is defined differently than the data in the target system.

Data Integration Distance: This is a metric that illustrates the level of difficulty involved in integrating (sharing) data and data services between two separate IT systems or Software Components. My testimony provides a framework for reducing Data Integration Distance among New Hampshire energy stakeholders. Later in my testimony, I will describe the considerable Data Integration Distance that exists today. Then, when I discuss Vision and Strategy, I will describe how to lessen that distance through use of an External Integration strategy using the proposed Data Platform.

Data Model and Logical Data Model: Logical Data Models defines a set of Data Entities, (e.g., accounts, sensors, tariffs, measurements) and their attributes (e.g., account number, meter ID,), their associations (e.g., readings taken from meters), and their relationships. A Data Model is not a physical database, but could be used to define such a database. Data Models can describe formats of data messages and/or packets transmitted between two separate IT systems. A Data Model is the agreed definition of data and is the key to Standardized Integrated Data Sharing I discuss later. Data Models can be documented and published and used by multiple parties. Please refer to the definition of "Common Data Model" I previously offered.

Data Platform and SB284 Data Platform: Dr. Jeffrey Taft, Chief Architect for Electric Grid Transformation at the Pacific Northwest National Laboratory, defines a platform as "an architectural concept of how elements are grouped, organized and related to each other" and also, as "a stable set of software components that provides fundamental or commonly-needed capabilities and services to a variable set of uses or applications through well-defined interoperable interfaces."[8] When the Legislature used the phrase "data platform" in the title of SB284 and the Commission likewise used the phrase in the caption of this docket, each was implicitly adopting this definition. In my testimony, Data Platform and SB284 Data Platform are synonymous and refer to my proposed platform. The Data Platform should incorporate platform architectural concepts defined by national experts. The proposed new IT infrastructure integrates siloed data, which is currently scattered inside legacy utility and non-utility Applications and databases, with External IT Applications that need that data to exist. The SB284 Data Platform should be a multi-use, Centralized, External Integration platform that is built around a flexible extendable Data Model designed both to support near term Use Cases immediately as well as to be extensible for support of potential future Use Cases such as those that exist in grid modernization and energy efficiency.[9] I am recommending the proposed Data Platform adhere to the NEEDS Standard (defined later) which will be designed to support Use Cases and External

[8] https://puco.ohio.gov/wps/wcm/connect/gov/38550a6d-78f5-4a9d-96e4-d2693f0920de/PUCO+Roadmap.pdf?MOD=AJPERES&CONVERT_TO=url&CACHEID=ROOTWORKSPACE.Z18_M1HGGIK0N0JO00QO9DDDDM3000-38550a6d-78f5-4a9d-96e4-d2693f0920de-nawqRqj

[9] A list of OCA proposed Use Cases can be found in Appendix 2 "OCA Suggested Use Cases with Brief Description".

IT Applications and Systems to comply with RSA 378:50. The Data Platform is illustrated in the center box (Box A) of Figure 1: “Platform Integration of Data Sources and Applications/Data Seekers.”

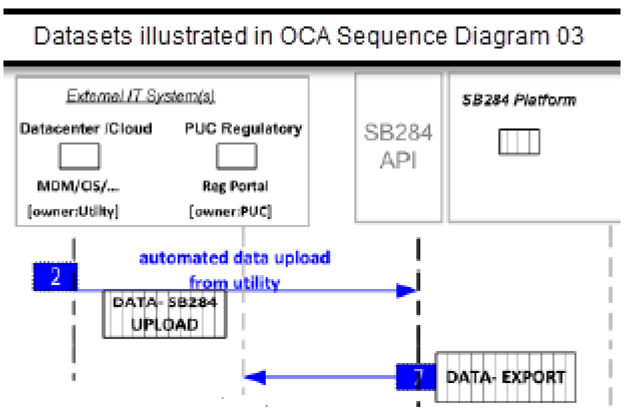
Data Seeker: A Data Seeker is a user or Application requesting data from the Data Platform. Data Seekers are shown in Figure 1 “Platform Integration of Data Sources and Applications/Data Seekers” in Box (C).

Data Sources: Data Sources include both utility data sources and non-utility data sources. Both of these sources are represented on Figure 1, “Platform Integration of Data Sources and Applications/Data Seekers” in Box (B) “DATA SOURCES.” Both utility and non-utility Data Sources should be included in the NEEDS Data Model.

Data Types and Data Categories: While not standardized or universally agreed to, these terms are often used in stakeholder discussion and proposals, to avoid the complexity of the Data Platform and underlying technical Data Models. Data Types, such as “customer,” “system,” “DER,” and “market/transaction” are discussed in the section on Vision. As an example, when I discuss the Logical Data Model, I use a Figure that shows categories named “Location,” Policy,” Technology/ Connectivity,” and “Energy Data” instead of discussing the more than 35 Data Entities represented by those Categories.

Dataset: Dataset is a generic term that refers to data records being uploaded to the Data Platform or downloaded from the Data Platform via the SB284 API. Datasets, or messages, exported from the SB284 Data Platform will conform to the agreed-upon and approved Data Model. Datasets can be visualized as spreadsheet data organized in columns and rows with each row called a record. However datasets can have hierarchy, such as an XML file. Datasets can also be visualized as in-transit messages within a Service Oriented Architecture. Datasets are depicted in OCA Sequence Diagrams (as the rectangular boxes with vertical lines that have arrows leading into and out of the SB284 API) and discussed in my Section on Vision.

Figure 2: Datasets Represent Data Uploaded to SB284 and Data Shared (Exported) by SB284 via Centralized SB284 API



Endpoint: Endpoints are physical hardware (often with internal software and data communication) connected to the power grid with capabilities of sensing or measuring power and energy at that location point. Endpoints have locations and are associated with physical real world assets that are owned by individuals or non-individuals (organizations).

ESPI: ESPI (Energy Service Provider Interface) is the basis of Green Button standards and specifications. The purpose is to "create[s] a standardized process and interface for the exchange of a retail customer's energy usage information between their designated data custodian (i.e.

distribution company) and an authorized third party service provider."[10]

Energy Data: Energy Data includes measurements of energy such as kilowatts and kilowatt-hours, and it also includes contextual energy-related data such as "time," "location," "tariff," etc. Shared data should be stored in databases that conform to the definition of the Data Model used to create the physical database. Data can also be contained inside messages from outside Applications or APIs requesting services from the Data Platform and can result from or effectuate an Application input and /or Application output.

External IT Applications / External IT Systems: These are Applications that exists outside of the Data Platform but are enabled by the Data Platform because they are provided authorized access to data through the SB284 API of the SB284 Platform. External IT Applications are discussed in the Section on Vision. "External IT Systems" is the term used in the OCA Sequence Diagrams (as one of the actors along the top bar) for many of the Use Cases and is considered synonymous with External IT Applications. External IT Applications and Systems are created by third parties but coordinated with the SB284 project team to ensure the SB284 API meets the data requirements. I provide a list of potential External IT Applications in the OCA Use Case document in Appendix 2.

External Integration: This means integrating data outside of the original IT System of the Data Source. In general, the term External Integration is a well-known and effective concept used to integrate systems that were originally not designed to be integrated and are de facto not

[10] https://openei.org/wiki/Green_Button_Developer

interoperable. For example, many legacy stand-alone proprietary back office systems at utilities are not interoperable with other systems. The term "external" means outside of the problem domain (outside of the legacy non-interoperable systems). In my testimony, External Integration means integrating the data that currently resides within each of the diverse sets of back office systems across each of the New Hampshire investor-owned utilities (IOUs) (as well as the customer-owned New Hampshire Electric Cooperative), as well as, potentially other Data Sources (such as competitive suppliers and vendors) whose data can be added to a separate centralized external system controlled by an agreed Logical Data Model such as NEEDS. This External Integration system is the proposed SB284 Data Platform. Once Data is combined and integrated in the SB284 Data Platform all utility data will exist at a statewide level, based on a Common Data Model (NEEDS). Because the Common Data Model will be published and documented, third parties will be able to understand and use the Centralized Data Platform with the appropriate permissions, and new External IT Applications can integrate with the Centralized Data Platform as a better alternative to going directly to each of the original Data Sources. External Integration enables External IT Applications by supplying robust standardized data not available today. In the Figure 1 diagram I presented before, the External Integration occurs with the Data Seekers box on the right where data outside of the original Data Source is integrated in the Data Platform and then provided to the Data Seeker.

Interoperability: This terms is based on the presence of data integration and the free flow of data and service among systems with low Data Integration Distance. Interoperability was the foundational goal established by the Department of Energy (DOE) and the National Institute of Standards and Technology (NIST) in the formation of the Smart Grid Interoperability Panel

(SGIP). As I discuss further in Section II, it is the over-arching justification for establishing Data Sharing standards. Interoperability is represented in the OCA Sequence Diagrams as blue arrows.

Logical Data Model means a Data Model as defined previously. There is no distinction between a Logical Data Model as used in RSA 378:51 I.(d) and a Data Model. "Logical" refers to the nature of the model as conceptual not physical. From the Logical Data Model one can create a physical database with software that conforms to the agreed upon Common Data Model. The Logical Data Model is used to create all the entities in the database such as the fields, the entries, unique IDs, and how data types are keyed to each other and indexed.

New Hampshire Electric and Gas Energy Data Standard (NEEDS): NEEDS refers to the Logical Data Model that I propose for development in connection with the SB284 Data Platform. Once the Logical Data Model is approved and therefore mandated, it becomes a "standard." NEEDS should leverage and be informed by the Logical Data Model work done by the DE 16-384 Data Working Group (DWG). NEEDS should facilitate an External Integration solution based on a Data Platform as proposed in my testimony. NEEDS defines the data inside the proposed SB284 Data Platform that third party Applications can understand and use. The NEEDS architecture is informed by local and national experts on utility operations and IT systems, a review of CIM and ESPI, a review of existing models and platforms deployed or under development in other US jurisdictions, and a review of data needs specific to New Hampshire law and Commission proceedings. NEEDS is designed as a future-proofing mechanism that should incorporate agreed-upon relevant parts of current and future data models plus internal New Hampshire requirements (based on agreed to long term Use Cases) in order to facilitate New Hampshire's statewide adoption of evolving national and international standards in coming years. NEEDS will leverage these existing standards and gear them toward New Hampshire. NEEDS is discussed further in the Data Platform section.

Relational Database Management System (RDMS): The RDMS is the set of rules that governs the data to conform to the Logical Data Model.

Service Oriented Architecture (SOA): This is the philosophy of encapsulating application logic in services with a uniformly defined interface (API) that can hide complexity and promote reuse. SOA is a design principle used in External Integration such as the proposed SB284 Data Platform.

Standardized Integrated Data Sharing (SIDS): SIDS is a form of data sharing in which data exchanged among separate but interoperable Applications and can therefore occur on an automated machine-to-machine basis. When one Application initiates a data access request directly to another Application, this is referred to as machine-to-machine interaction, with no human involved. Standardized Integrated Data Sharing facilitates machine-to-machine data access.

Statewide Index and Index: Statewide Index is a core architectural trait and function of the Data Platform's database and should be based on the NEEDS Data Model. All New Hampshire Data contained in the Data Platform is unique including location readings, meters/sensors, and transactions. Therefore, the data can be searched, updated, managed, reported and validated based on a statewide (not utility level) set of unique system IDs. Statewide Index is discussed further in the Data Platform section.

Use Cases: Use Cases typically define business requirements of an Application.[11] For purposes of my testimony, Use Cases are in one of two categories: Stakeholder Application Use Cases and a Master Use Case for the proposed Data Platform.[12] Use Cases are supported by and dependent on a properly designed Logical Data Model. Use Cases can be mapped to user stories (as described in the testimony of Ethan Goldman being filed on behalf of Clean Energy New Hampshire). Most of these Use Cases will be developed by third parties as External IT Applications provided the SB284 Platform is developed in a way that can expand for future cases.

II. History of Data Sharing in the U.S. and New Hampshire Problem Statement

A. Early national efforts to establish data standards and Green Button

Q. Briefly describe the history of standards-based energy data sharing around the country, relative to SB284 and customer data.

A. In a period roughly from the 1990s through the 2010s, as the U.S. electric grid was aging and demand for electricity was rising, significant major blackouts were occurring in metropolitan areas. New metering and monitoring systems were created with help from the Electric Power Research Institute (EPRI). EPRI is an independent non-profit started to support research and development to address technical challenges in the electricity sector. In early 2009 the National

[11] https://en.wikipedia.org/wiki/Use_case

[12] Appendix 2 contains the OCA's proposed Master Use Case for the development of the Data Platform and a number of Use Cases that are examples of potential stakeholder Applications.

Institute of Standards and Technology (NIST)[13] contracted EPRI to develop a Smart Grid Interoperability Standards Roadmap. In late 2009 NIST established the Smart Grid Interoperability Panel (SGIP) as a private/public partnership to define requirements for essential communication protocols.[14] The SGIP was compromised of energy and technology domain experts and the goal was to solve interdependency and interoperability issues with the US electric grid. The effort included analysis and development of standards for the electric grid.

At roughly the same time, and in part inspired by data standardization efforts of medical data (as well as the Blue-Button standard adopted for the military sector), Secretary Steven Chu of the U.S. Department of Energy decided a similar portability effort should occur for energy data. The goal was to promote innovation and provide increased transparency by allowing energy data to be downloaded by consumers (similar to what was becoming possible with medical data). This decision would lead to efforts to create a standardized process and interface for retail customer energy usage data. In the spirit of public-private collaboration, some key standards from SGIP and the Common Information Model (CIM) became the North American Energy Standards Board (NAESB) Energy Service Provider Interface (ESPI) standard now known as Green Button.

[13] "Under the Energy Independence and Security Act (EISA) of 2007, the National Institute of Standards and Technology (NIST) has "primary responsibility to coordinate development of a framework that includes protocols and model standards for information management to achieve interoperability of smart grid devices and systems…" [EISA Title XIII, Section 1305]" https://www.nist.gov/system/files/documents/smartgrid/Report_to_NIST_August10_2.pdf

[14] https://www.nist.gov/programs-projects/smart-grid-national-coordination/smart-grid-interoperability-panel-sgip

Q. Briefly describe the Green Button standards created in 2011.

A. The Green Button standard is based on elements of the broader Common Information Model (CIM). Green Button's scope is limited. Green Button is a customer-centric data sharing standard specifically designed to share customer usage data. Green Button is an industry standard that includes XML[15] data file format and API for accessing and securely sharing energy (electric and natural gas) and water usage data.

Green Button is an open data standard and is maintained by the Green Button Alliance (a non-profit organization founded in 2015). The Green Button standard is a well-known, consent-based, U.S. standard based on a model that places the customer at the center. Green Button is part of the DOE's DataGuard Partnership Program, which is an energy data privacy program.

Q. How is the Green Button Standard governed?

A. Changes and new requirements are addressed through the OpenADE Task Force.[16] Both members and non-members of Green Button Alliance can become involved in improving the standard using this mechanism.

Q. How widely is Green Button CMD standard used?

A. Today Green Button is deployed in at least seven states that have implemented a Connect My

[15] https://en.wikipedia.org/wiki/XML

[16] "The OpenADE Task Force defines systems requirements, policies and principles, best practices, and services, required for information exchange and control between 3rd Party energy usage data analysis providers (EUDAPs), utility enterprise front and back office systems, and consumer customers." http://osgug.ucaiug.org/sgsystems/OpenADE/default.aspx

Data (CMD) to allow automation and machine-to-machine data sharing for advanced applications.[17] In 2012, the Ontario Ministry of Energy initiated an effort to mandate Green Button for the entire Canadian province. Ontario utilities were scheduled to have fully adopted Green Button by July 2020. [18]

Q. What are Green Button's Use Cases?

A. Green Button itself has two Use Cases: "Download My Data" and "Connect My Data" (CMD). CMD provides integrated access to software Applications.

Q. Does Green Button's Connect My Data Use Case represent Standardized Integrated Data Sharing as you have defined earlier?

A. Yes, the Green Button Data Model and API facilitate interoperable machine to machine data sharing process.

Q. How has the Green Button standard evolved during this period of grid modernization?

A. The Green Button standard has evolved as a secure model that can support well designed successful vendor-built Green Button platforms. The Green Button standard and API have evolved in recent years including the areas of privacy/consent, [19] authorization, and security. The Green Button standard has also evolved in response to active industry comment and requests for enhancements. Today, vendors implementing the Green Button standard are also required to

[17] *See* Misson:Data, "Energy Data Portability" (2019) (available at http://www.missiondata.io/) at 6 (noting that Green Button CMD had been implemented or would be implemented in 2000 by utilities in California, Colorado, Illinois, Michigan, New Jersey, New York, and Texas).

[18] *See* https://goenergy.ca/ontario-utilities-implement-green-button-data-standard-july-2020/.

[19] https://www.greenbuttonalliance.org/index.php?option=com_dailyplanetblog&view=entry&category=technical&id=20:green-button-ensures-utilities-and-third-parties-protect-customer-privacy

receive certification from Green Button Alliance. There are increasing examples of technology vendors implementing the Green Button standard and API to deliver secure, customer focused (customer friendly user interfaces) experiences that allow customers to share their data, with consent, in a user friendly approach.

Q. Is Green Button part of your recommendation?

A. Yes, Green Button Connect My Data is included in my recommendation and is also required by RSA 378:53.

B. Evolving Data Sharing Needs from 2020 Onward

Q. Have the needs for standards based data sharing evolved beyond what Green Button envisions?

A. Yes, the depth and scope for data sharing is increasing for a variety of often overlapping reasons. There is an expanded, often common, data sharing requirement that exists across three domains discussed below: grid modernization, clean energy, and the Industrial/Residential Internet of Things (IOT). As I will discuss shortly, there are many Use Cases across these three areas that work with common underlying core data. Grid modernization efforts are ramping up and changing the electric grid in fundamental ways. Data is center stage in grid modernization because it is a necessary component of success and is needed to capitalize on the significant financial investment smart grid requires.

Environmental goals, sustainable energy goals, and "prosumer"[20] goals are also driving

[20] "Prosumers" are electric consumers who also produce or provide their own electricity with renewable energy and storage devices.

needs to access data and analyze usage data by location and demographic or policy groups in order to develop clean energy services and solutions.

There is a digital transformation of industry and IOT. Advances in technology are expanding a system of interconnected devices creating opportunity for industry, consumers and prosumers. Broadly referred to as IOT or IIOT (Industrial Internet of Things) and IOT Smart Cities,[21] these three Use Cases produce and also consume Energy Data.

Q. How does grid modernization increase the scope of data sharing?

A. Grid modernization is creating new Data Models and Data Types integrated into the economy using APIs to serve innovative third party Applications and platforms. Based on my research and outreach, grid modernization both creates data and, at the same time, requires the emergence of Applications that are heavily, or often entirely, dependent on accessing the rich granular data generated by new grid modernization assets.

Q. What significant challenges will grid modernization pose that need to be considered in DE 19-197?

A. Grid modernization will present three significant challenges to be considered in this docket related to quantities of data, new data science, and the economy.

Grid modernization creates an enormous amount of energy data because of the interaction of more devices, more granular measurements, more regulated and unregulated sources, and

[21] "Smart cities use Internet of Things (IoT) devices such as connected sensors, lights, and meters to collect and analyze data. The cities then use this data to improve infrastructure, public utilities and services, and more." https://www.businessinsider.com/iot-smart-city-technology

more third party products and services creating data. There will be new types of data, new data elements, new ways data can be related given modern capabilities of the grid.

The data science of extracting meaningful information that is useful for data sharing is different from the science of designing traditional utility back-office systems. The design and methods of extracting meaningful information from large amounts of complex data generated by Smart Grid for sharing are intended to result in Standardized Integrated Data Sharing for Stakeholder Applications such as in the OCA's Suggested Use Cases in Appendix 2. Utility back-office systems are designed to run the utility's business (collect and validate AMI readings, produce customer bills, track grid asset geographically, produce financial data, etc.) and are not designed to provide Standardized Integrated Data Sharing.

Beyond just the energy sector, the overall economy is actually now dependent on data Interoperability. Integrated data is the goal. Robust integrated energy data sources are needed to support new industries and energy based economies both in New Hampshire and nationally. Current utility systems were never designed to handle the challenge of integrating with external Applications. They share Energy Data using technologies optimized for human consumption (PDFs, excel, printed reports and statements, and undefined or inconsistently defined datasets) when providing information requested by external users. However, in a connected economy, Applications, not humans, need to access standardized Energy Data directly through APIs (referred to as machine to machine). Because there is currently a lack of standardized Energy Data accessible through APIs, integrating the data from these Data Sources is nearly impossible. Where integration is possible through manual intervention, the costs are prohibitively high and create barriers to market development and transformation. With a Data Platform as proposed, New Hampshire Energy Data Sources could migrate toward a model of Standardized Integrated Data Sharing (defined previously) as a fully interoperable form of data sharing to facilitate a competitive energy market that stimulates innovation and the economy. A blueprint for a targeted migration to robust Energy Data sharing is shown in Figure 11(Proposed Build) later in my testimony.

Q. Have economic, environmental, and sustainability goals increased the scope of data sharing?

A. Yes, there is a growing confluence of synergistic goals in the areas of clean energy, energy efficiency, and prosumer empowerment that could have transformational economic and environmental benefit to New Hampshire. As a group, the multitude of projects, public/private business initiatives, investments, and policies required to realize goals, all either need or would greatly benefit from access to electronically organized and standardized data.

Q. Could you provide an example of a resource that supports the concept you just described?

A. A policy paper by the U.C. Berkeley School of Law in California, part of a series of reports on how climate change creates business opportunity, discusses a broad range of the Data Types that are "immensely useful to a variety of audiences" to "realize both economic and environment benefits."[22] The paper bifurcates Energy Data into two main data categories: "utility centered data" and "customer centered data". The utility-centered data includes distribution infrastructure data, transmission infrastructure data, and aggregated data reflecting consumer behavior. The

[22] Knowledge is Power https://www.law.berkeley.edu/files/Knowledge_Is_Power.pdf.

customer-centered data includes data from meters, Internet of Things, distributed energy resources (DERs), tariffs, energy efficiency policies and customer segmentation data.

The Berkeley Law report states:

> Ultimately, expanding access to energy data can bring cleaner, more efficient power to Californians and help them save money. It could also boost emerging clean technologies, such as renewable energy, energy storage, electric vehicles, demand response, and energy efficiency measures, which will help the state meet its environmental goals in a more cost-effective manner. And all ratepayers could save money as improved energy efficiency decreases the need for new investments in polluting power plants.[23]

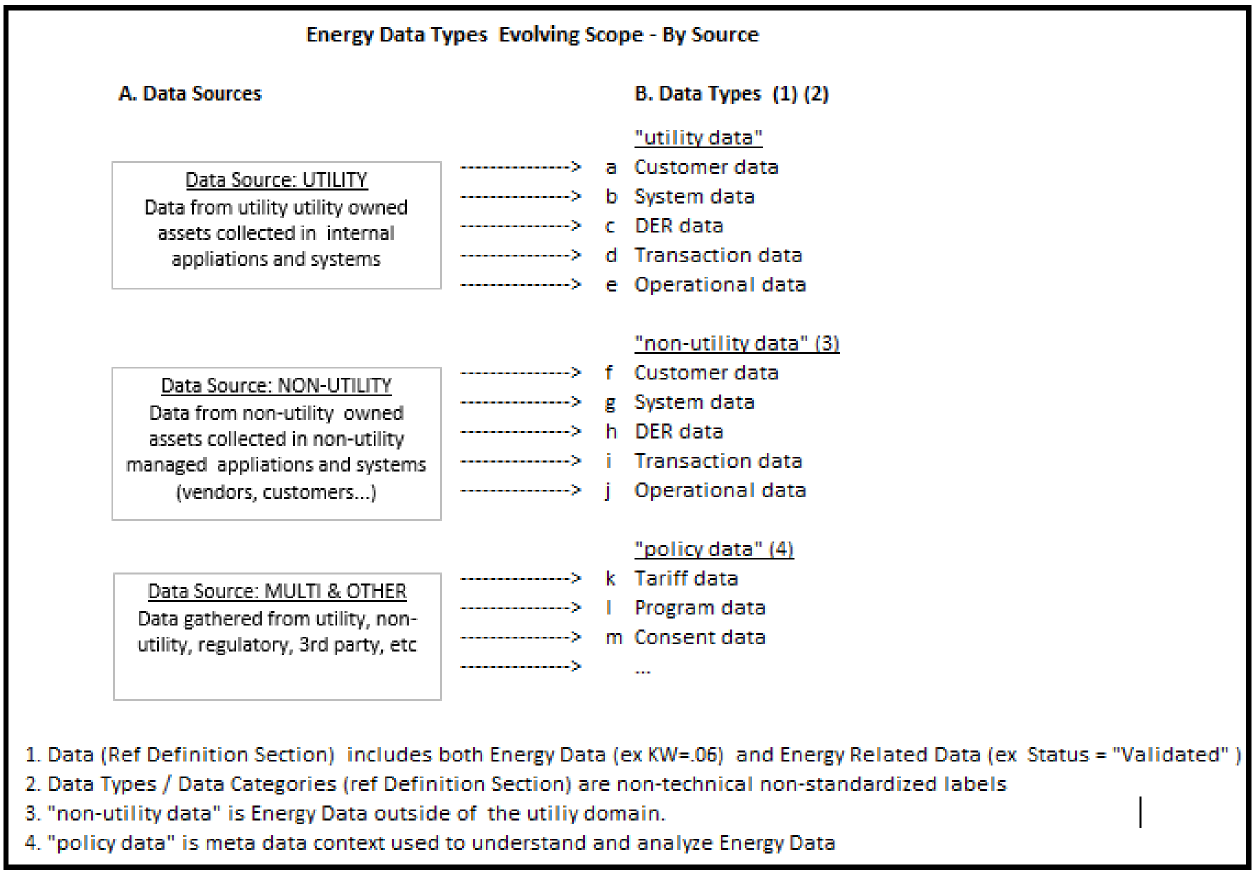
Q. How are data sharing needs evolving in 2020 and beyond?

A. Looking forward, the scope of data needed for sharing is expanding in a number of ways in order to support grid modernization, clean energy, and the IOT digital transition. I have provided Figure 3, "2020 Forward Looking Data Sharing," as a visual representation. First, utility data needs to be shared in different ways than traditionally by engaging in external data integration. Second, we need to move beyond just customer data to include more Data Types (see Column B in Figure 3) that could be increasingly used by modern applications such as community energy dashboards, community aggregation, etc. Third, we need more data from more Data Sources (see Column A in Figure 3) to be included and integrated such as from prosumers, delivered fuel companies and other energy service providers. The Data Platform should be structured in a way that eventually allows these non-utility sources to upload their data so it can be shared with various Applications. The Data Types and the three kinds of Data Sources represented in Figure

[23] *Id.* at 1.

3 are important parts of the NEEDS Logical Data Model. Including a provision for these forward looking Data Types can provide a Data Platform with future-proofing aspects by making it extensible to support future Use Cases.

Figure 3: 2020 Forward Looking Data Sharing



Q. How does IOT increase the scope of data sharing?

A. The Internet of Things generates Energy Data and consumes Energy Data associated with the distribution grid. IOT, which is a system of interrelated computing devices or mini Applications with unique identifiers, results in a variety of energy-related Use Cases. For example, sensors and networked devices performing energy management in buildings is a classic IOT Use Case. IOT also has smart city Use Cases where sensors and networked devices (utility and non-utility owned) are deployed on utility poles, street lights, and strategic locations. These sensors and devices have the capability to monitor and/or manage energy loads for communities down to the circuit level or sensor level. This capability can facilitate collecting data needed for the analysis of improvements on the system or the evaluation of demand side resources.

Q. Is there an overlapping common set of data that is used or needed by the three domains (grid modernization, clean energy and IOT) discussed previously?

A. Yes. Fortunately, there is considerable overlap in the data that is needed for various Use Cases and Applications. Based on research and outreach with energy experts, grid modernization experts, and other stakeholders nationally the OCA has outlined the Data Types necessary to build various types of Applications in these domains. We recommend that this work be used as a basis for the development and testing of the Common Logical Data Model. Figure 3 that I provided earlier outlines the common Data Types required for these Applications in these domains to meet energy and environmental goals. In the Data Platform section, I present Figure 9, "4 Categories of Data in NEEDS," to provide additional detail on how those Data Types relate generically and logically to act as a foundation to a statewide Data Platform.

Q. Can Green Button be used to share data for Data Types shown in Figure 3?

A. Green Button is not currently a solution for all the Data Types shown in Figure 3.

Q. Do frameworks and national standards exist that incorporate all of the Data Types in Figure 3 and facilitate Standardized Integrated Data Sharing?

A. Yes and no. It is important to note that many groups worldwide are working toward development of new models that advance Standardized Integrated Data Sharing. Ideas are shared, reused and improved upon. A possible scenario for the SB284 Data Platform will be models that combine existing US and international standards, or aspects of those standards with additional ideas, which is not unlike how the DE 16-384 DWG Logical Data Model began because it was influenced by Green Button. The proposed continuation of the development of a New Hampshire Logical Data Model (NEEDS) would continue the process of evaluating other standards as part of that model. New models should support electronic sharing of categories of advanced Use Cases such as those contained in Appendix 2, or the Use Cases submitted by other parties in this proceeding, or Use Cases being piloted nationally.

Q. Have other states recognized the need for standardized electronic access to more Data Types?

A. Yes. Data sharing efforts have begun in many other states including California and New York, which I discuss in the next section. Data sharing has been driven by competitive markets, increased data availability due to modern meters such as AMI, increased ability to process and store vast quantities of data, development of technological solutions to energy needs, and carbon reduction.

C. Data Sharing Influence from Utilities and Regulators in Other States

Q. What outreach have you performed in other states and provinces relative to their data sharing efforts?

A. I have had, to varying degrees, discussions on data sharing systems in other states with regulatory staff and or utility staff with direct knowledge of platforms in those states. In some cases sharing of ideas included technical documentation, technical citations, and frank discussion of strategies, problems, and visions for that state. New Hampshire's proposed SB284 platform was also frequently discussed. Some of my state outreach is summarized as follows:

In California I had discussions with the Pacific Gas & Electric staff involved in the Green Button Platforms. In addition, California PUC Staff provided technical resources on data sharing and their states implementation of Green Button standard on their website.

I had discussions regarding the Smart Meter Texas (SMT) including discussion with staff from the various utilities that upload to SMT including Oncore, TNMP, CNP and AEP to review the history and functionality and design of the system.[24]

I had discussions with Chief of Grid Modernization and Security of the Ohio utility commission (PUCO) as well as the analysts who coordinated Power Forward, PUCO's effort to develop the road map to Ohio's electricity future.[25] In addition, I participated in PUCO's Data and Modern Grid Workshop facilitated by EnerNex.

Regarding Illinois, I had discussions with data experts working on The Big Energy Data Center platform and conducting research based on data collected by the platform.[26]

I had discussion with managers, technical staff, and Green Button members from Ontario regarding London Hydro's development of a centralized energy data and Green Button

[24] Issues with the SMT system are discussed by MissionData: http://www.missiondata.io/news/2017/9/22/5-things-you-dont-know-about-smart-meter-texas.

[25] The effort was conducted over three phases and included 127 experts and approximately 100 hours of education. This nationally recognized roadmap can be reviewed in the report available here: https://puco.ohio.gov/wps/wcm/connect/gov/38550a6d-78f5-4a9d-96e4-d2693f0920de/PUCO+Roadmap.pdf?MOD=AJPERES&CONVERT_TO=url&CACHEID=ROOTWORKSPACE.Z18_M1HGGIK0N0JO00QO9DDDDM3000-38550a6d-78f5-4a9d-96e4-d2693f0920de-nawqRqj.

[26] https://www.citizensutilityboard.org/welcome-big-energy-data-center/.

platform.[27]

I had discussions with the data team at the New York Department of Public Service regarding the development of a New York DER pilot which will follow a centralized model. In addition, I had discussions with the Director of the Rhode Island National Grid Advanced Metering Infrastructure (AMI) division. I also had technical and business discussions with ConEd at their headquarters in New York City regarding their AMI Green Button platform.

In Massachusetts I had discussions with the Program Manager of National Grid's Transformation Office and a senior data analyst regarding aspects of data models being developed in National Grid relative to the efforts in New Hampshire.

In Vermont I had discussions with the Vermont Energy Investment Corporation (VEIC) regarding the Vermont Community Energy Dashboard as a model for New Hampshire municipalities to use the data that will become available as a result of the SB284 data Platform.[28]

Q. What data efforts outside New Hampshire have significantly influenced you relative to your recommendation for a centralized Data Platform?

A. I have been influenced by both what has and has not worked in other jurisdictions. In a number of ways California has led data sharing efforts in the US including mandating all IOUs to share data, on a utility by utility basis in 2014.[29] Efforts in California have been insightful from the

[27] https://www.londonhydro.com/site/binaries/content/assets/lhcontent/news/mediarelease_lhteamfinalistpowerforwardchallenge.pdf.

[28] https://www.vtenergydashboard.org/about.

[29] On May 1, 2014, the California PUC issued Decision 14-05-016 http://docs.cpuc.ca.gov/PublishedDocs/Published/G000/M090/K845/90845985.PDF adopting energy data access rules including a decision to implement data sharing systems on a utility by utility model

perspective that California is seven years into a decentralized (utility by utility approach) Green Button data sharing effort. My analysis has led me to the opinion that ongoing data quality issues that exist across California's IOUs are the result of issues that exist in a decentralized model. Based on this, and analysis in other areas and states, my recommendation for New Hampshire is to pursue a centralized data sharing model. This centralized model is also a strategy that is being pursued in New York in an ongoing DER pilot. My understanding, based on my outreach discussed earlier, is that the pilot is testing a data model and platform. One of the data requirements of the pilot includes the ability to enable complex, developer-designed, select queries across all categories of customer. There is a strategy to consolidate the data of all New York utilities into a centralized version of the DER platform being piloted.

Q. Please give an example of "data quality issues" relative to data sharing.

A. Data issues can include inconsistent data. A white paper from Mission:Data lists "incorrect data" as a performance problem in utility IT systems. One of the types of performance problems listed is "[i]ncorrect data are data sent to a third party that do not match what the customer sees on the utility's portal."[30]

Q. What evidence exists that leads you to the conclusion that "data quality issues" may be associated with a decentralized data sharing model?

therefore rejecting a an alternative centralized data platform model presented in a September 2012 CPUC Briefing Paper titled "Energy Data Center."

[30] https://static1.squarespace.com/static/52d5c817e4b062861277ea97/t/5c3a849b562fa75d70fd7953/1547338949271/Energy+Data+Portability.pdf at 9.

A. Data quality issues can cause unpredictable results. In California's decentralized Green Button model, third parties using the utilities platforms experience inconsistent unpredictable results. That has led to vendors resorting to custom integration work. For example, a user of one of the California utility Green Button platforms reported: "We have separate code for each California utility. Their implementations are totally different from one another."[31] Also, I am aware of inconsistent performance between the separate utility Green Button platforms. In a complaint before the California PUC, a vendor alleges significant differences in performance between PGE Green Button implementation and SDG&E's Green Button implementation. I have anecdotal information that data quality issues have been a particular problem in situations where there are multiple customer meters.

Q. Is it your opinion that poor system design by the utilities is the cause of data quality issues?

A. No. It is my opinion that strategically, the decision to mandate a utility by utility data sharing model seven plus years ago was, in hindsight, the larger issue. Attempting to produce an identical IT result from separate IT systems (separate IOU datacenters) is very difficult, even if all agree to use the same standard, such as Green Button. Separate technologies, separate testing environments, separate SDLC's managed differently, in totality, are unlikely to produce identical results.

Q. What is your conclusion on centralized vs decentralized data sharing models?

A. In my opinion, poor data quality results from decentralized systems and is the number one risk to

[31] *Id*. at 12.

harm a data sharing effort. In addition, it is my opinion that inconsistent privacy models can emerge in a decentralized model. The issues of inconsistent data sets between utilities, different API implementation, are largely reduced or eliminated using a centralized model. My recommendation is to pursue an External Integration strategy using a centralized SB284 Data Platform as the means to implement that model. This recommendation will be further illustrated later in Figure 11 (Proposed Build).

D. New Hampshire Status and Problem Statement

Q. Does New Hampshire currently have Standardized Integrated Data Sharing?

A. There is no existing Standardized Integrated Data Sharing for Energy Data in New Hampshire. The vast majority of utility Energy Data today is blocked off or inaccessible to external Data Seekers because it resides inside the many legacy proprietary back office IT systems run by the utilities. In the realm of interoperability and data sharing, this data is said to reside inside a data silo. There is some level of data access, mostly human-to-human or human-to-machine, which is provided through utility web sites, the NH Saves energy efficiency website, Eversource's Customer Engagement Platform, and varying degrees of implementations of Green Button Download My Data also available on utility web sites. I do not consider this Standardized Integrated Data Sharing because it does not support machine-to-machine API data access that innovative third party Applications and systems require.

Q. What is meant by data silo?

A. The terms “silo,” “software silos,” and “software stovepipes” are used in EPRI’s “An Introduction to the CIM for Integrating Distribution Applications and Systems.” Specifically, the EPRI paper states:

> In the typical distribution utility there are hundreds and even in some cases thousands of software solutions and applications that are managed by the IT department. These applications were initially used and operated independently by the various groups, departments, and organizations in the utility. Whenever a business process required information to flow from one system or application to another system or application, the process was managed manually. This level of maturity in an organization's software infrastructure is sometimes called "islands of automation," "software silos," or "software stovepipes."[32]

For purposes of my testimony, I define "data silo" on two levels with cumulative effects. At the first level is an Application that creates or generates Energy Data for its internal use (the Use Case or specialized purpose for which that Application was intended), but does not support Standardized Integrated Data Sharing with other Applications. Such an Application can be referred to as a data silo because the data only exists within that application and cannot be shared. The second level, adding significantly to the data interoperability challenges, is the fact that each IOU is in itself a silo covering its franchise territory. IOU data systems are not built to integrate with external systems, including those of neighboring IOUs. This is further exacerbated by the fact that the IOU systems and IT processes, are all planned, designed built and operated differently and use different IT technologies and data models. The result is an entrenched roadblock to Standardized Integrated Data Sharing between statewide utilities and the emerging new energy economy that needs access to that data.

Q. Since this docket deals with sharing Energy Data, not integrating utility distribution systems, what is the relevance of software silos in utilities to SB284?

[32] EPRI Common Information Model for Distribution, http://www.tut.fi/eee/research/adine/materiaalit/Active%20network/ICT/EPRI%20CIM%20for%20distribution.pdf.

A. Utility distribution Applications and systems contain a major amount of the data that should be shared for a "wide range of applications and business uses."[33] In other words, utility systems and applications prevent access to Energy Data which inhibits robust data sharing in New Hampshire. The data, inside the utility software silos, must be placed in an environment that supports Standardized Integrated Data Sharing. That occurs by migrating toward a Common Logical Data Model using the SB284 Data Platform.

Q. How is energy data shared in New Hampshire currently?

A. In New Hampshire, data sharing currently occurs in a couple of ways. Utility customers generally receive their data through printed or on-line monthly customer bills. Some receive data through Home Energy Reports via their utility's ratepayer-funded energy efficiency program or, in the case of Eversource, through that utility's Customer Engagement Platform. In some utility territories, customers can access their data through the Download My Data platform and provide that file to third parties. However, this is limited to the information from one utility and is not directly shared with third parties. This is not the preferred Green Button Model, called Connect My Data. Connect My Data allows modern External Applications to directly access the customer's usage data with consent. Data is also shared using the Electronic Data Interchange (EDI). This is a very old standard that was initiated by the New Hampshire PUC in 1998.[34] According to a statement made by Eversource during one of the technical sessions in this proceeding, EDI does not support TOU or interval data. It is also my understanding that EDI

[33] RSA 378:51 I(a).

[34] https://www.puc.nh.gov/Electric/edi.htm.

does not have a published API that supports Standardized Integrated Data Sharing as defined in my testimony. Ad hoc data sharing requests from customers are frequently handled manually and in non-repeatable processes and with unpredictable results.

Non repeatable IT processes are referred to as having low levels of maturity by Carnegie Mellon University Software Engineering Institute's Capability Maturity Model (CMM). The Software Engineering Institute has defined five CMM levels of software process maturity.[35] From a Standardized Integrated Data Sharing perspective, New Hampshire's data sharing would be described as immature under the CMM. This is my informal opinion and not based on SEI analysis or input. The point is there is a lack of repeatability in New Hampshire's data sharing processes. The low level of data sharing maturity reflects non-formalized or poorly defined data sharing processes that do not produce "predictable" results found in organizations with higher levels of maturity. As noted by Dr. Amro Farid of the Thayer School of Engineering at Dartmouth College, at the technical session following the prehearing conference in this docket, Data Seekers in New Hampshire can wait weeks and months before receiving a dataset. Data Sharing is also frequently accompanied by significant operational inefficiency, systemic recurring manual work, poorly defined procedures (non-repeatable), and data 'work arounds' (performing a business tasks or an IT process while not having the data that ideally is required for that task). These inefficiencies hamper both the Data Sources (often utilities) and the Data Seekers (usually customers, municipalities, and third party energy providers, etc.).

[35] SEI's "The Capability Maturity Model Guideline for Improving the Software Process" defines five levels of maturity: CMM1 "Initial" CMM1, "Repeatable" CMM2, "Defined (Standard) CMM3, "Managed ("predictable) CMM4 and 5) Optimized CMM5. See https://resources.sei.cmu.edu/asset_files/TechnicalReport/1993_005_001_16211.pdf.

Q. Please provide a few example of data sharing inefficiencies in New Hampshire.

A. Examples include manual work performed by utility customer's energy managers who allocate staff to perform manual tasks such as typing data from energy bills (multiple page bills in the case of larger organizations or state agencies) into local databases (Applications). There is significant use of non-secure screen scraping[36] of data as another approach to this same problem. There are numerous duplicative efforts by municipalities to set up small consulting/project teams to formulate requests for energy data from multiple utilities, incurring time and costs. The information then comes from the utilities in excel or PDF format that is different from utility to utility or lacking in detail. Third party vendors hoping to offer a potential customer a proposal for energy solutions (for example, a proposal to install distributed energy resources for sale or lease) frequently incur significant manual effort interpreting customer energy bills.

In other Use Cases, large datasets of Energy Data are shared in Excel files, where design of the Excel file is not standardized or well documented, and therefore can require further human-to-machine manual input on the part of the Data Seeker to get the data out of the Excel file and its Application. Statewide there is a small army of people tasked with getting required Energy Data, stored at utilities, into their own databases. Throughout New Hampshire energy managers in private organizations and state agencies maintain in-house staff or FTEs spending time manually entering usage data and information on multi-page utility bills, typing that information into a Databases so that their Applications can analyze the usage, verify billing

[36] Screen scraping is the process of collecting screen display data (for example the html on a web page) so that the scraped (copied) data can be used by another Application such as a database. Security and authentication credentials are sometime shared depending on the website being scraped (for example a customer login area of their utility website where account number and password are required). Data quality issues are not uncommon using this technology.

accuracy, and to a limited degree analyze energy efficiency in each of the buildings. Finally, and very significantly, these manual data processes have led to proliferations of inefficient standard operating procedures across the state in organizations because processes must be built around (and limited by) systemic inefficiencies surrounding data access. The inefficiencies themselves are not the problem. Rather, they are symptoms of the problem that should be solved in this docket.

Q. What is the underlying problem that causes the symptoms of manual effort, compromised practices and procedures, immature processes, and lengthy waiting periods?

A. There are two fundamental problems blocking efficient and effective data sharing in New Hampshire: (1) poor data quality, and (2) lack of electronic access to data.

1. Poor Data Quality: lack of good data. Specifically, there is an absence of statewide datasets of clean, understandable, and usable (programmable) data. External IT Applications and systems can be the solution to elimination of the inefficiencies and manual work. However, Applications require precise definitions of the data they consume or they will operate ineffectively, produce data errors or missing data, and create other software bugs. Software developers, hired by stakeholders to create standards based Applications, rely on documentation and definitions contained in models, such as Logical Data Models, that are specified by the stakeholder. A tested, well defined Logical Data Model, designed to support the Use Cases of the organization, can help reduce chances of data errors (related to poor data definitions). Applications also need access to all the data required for their Use Case. In many cases, this means data from more than a single utility franchise. The absence of good data that conforms to definitions in a robust Data Model blocks Applications from being built. Therefore the first problem I identify as causing

symptoms of data sharing inefficiencies in New Hampshire is the lack of good data.

2. Lack of electronic access to energy data. Applications that automate processes and eliminate manual work are designed to access data electronically on their own, machine to machine. Today, New Hampshire data sharing primarily involves manual access to data through the transmission of excel and PDF files. This introduces delay, creates errors, and precludes program access by an Application. Manual access includes human-to-human and human-to-machine. Data is often requested manually by phone, email, written discovery requests, or part of a consulting engagement dedicated to acquiring Data. In general, data dumped to Excel or .pdf is then manually delivered as e-mail attachments, thumb drives, or other methods not usable by advanced Applications. The lack of electronic access blocks or vastly increases the costs for (and therefore viability of) the development of new Applications that can solve and automate energy, grid, and climate problems. Therefore, the second problem I identify as causing symptoms of data sharing inefficiencies is the lack of automated electronic access to data.

These two problems, poor data quality and lack of automated electronic access, combine to create significant but invisible friction throughout New Hampshire's energy environment. This friction leads to systemic inefficiencies that create costs and limit economic growth.

Q. Are these two problems, which you refer to as friction, unique to New Hampshire?

A. No, they are very common nationally.

E. Data Integration Distance and Economic Friction in New Hampshire

Q. Do you suggest a single metric that measures the friction and the extent of both road blocks caused by poor data quality and lack of access to data?

A. Yes. Data Integration Distance, which I previously described in the definitions section of my testimony, would be an appropriate visual metric for New Hampshire. The term "Integration Distance" is used in the EPRI technical paper entitled "The Common Information Model for Distribution."[37] The focus of the article is systems integration, of which data is a core aspect. I have modified EPRI's term, by adding the word "data" to focus specifically on the Integration of Data, which is the problem being addressed in SB284. While "distance" in this context is subjective, it nevertheless represents time, cost, friction and brick walls that organizations encounter and attempt to overcome in efforts to obtain quality Energy Data.

Q. Please illustrate the Data Integration Distance metric visually.

A. I have created Figure 4, "Data Integration Distance – concept," showing the spectrum of data interoperability. It can be thought of as a visible indicator of friction in the energy economy that is otherwise invisible.

The upper portion of the diagram shows the desired outcome, full data integration. In an ideal scenario, Data Source A receives a request and provides data to the target Application within seconds[38] of the request in which Data Source A seamlessly consumes and uses the data with no added custom integration cost. This scenario results in no gap or zero Data Integration Distance.

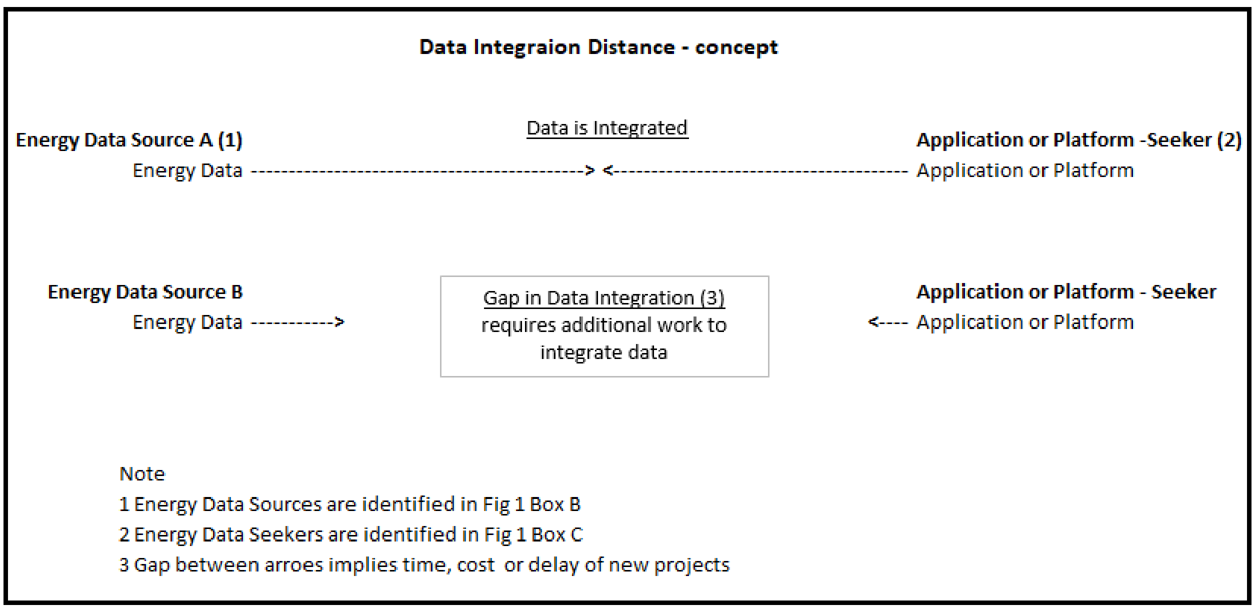
The lower portion of the diagram shows the high-friction, problem scenario with poor

[37] EPRI Common Information Model for Distribution
http://www.tut.fi/eee/research/adine/materiaalit/Active%20network/ICT/EPRI%20CIM%20for%20distribution.pdf.

[38] Large data sets, such as research datasets will take longer to process (minutes, hours) and longer to deliver depending on streaming and network capability.

data integration. In this problem scenario, Data Source B receives a request, must manually access and generate the data into a format it can share, and then provides data that is effectively unusable by the target Application such that the Data Seeker has to perform additional work for their Application to use the data. This could happen for any number of reasons and requires custom integration resulting in more Data Integration Distance, and added additional costs to data sharing. The greater the gap (Data Integration Distance) the greater the time and cost to the project and the New Hampshire economy.

Figure 4: Data Integration Distance – Concept



Q. What causes Data Integration Distance?

A. As I said, the major drivers of the gap in Data Integration Distance and non-interoperability are poor data quality and lack of electronic access.

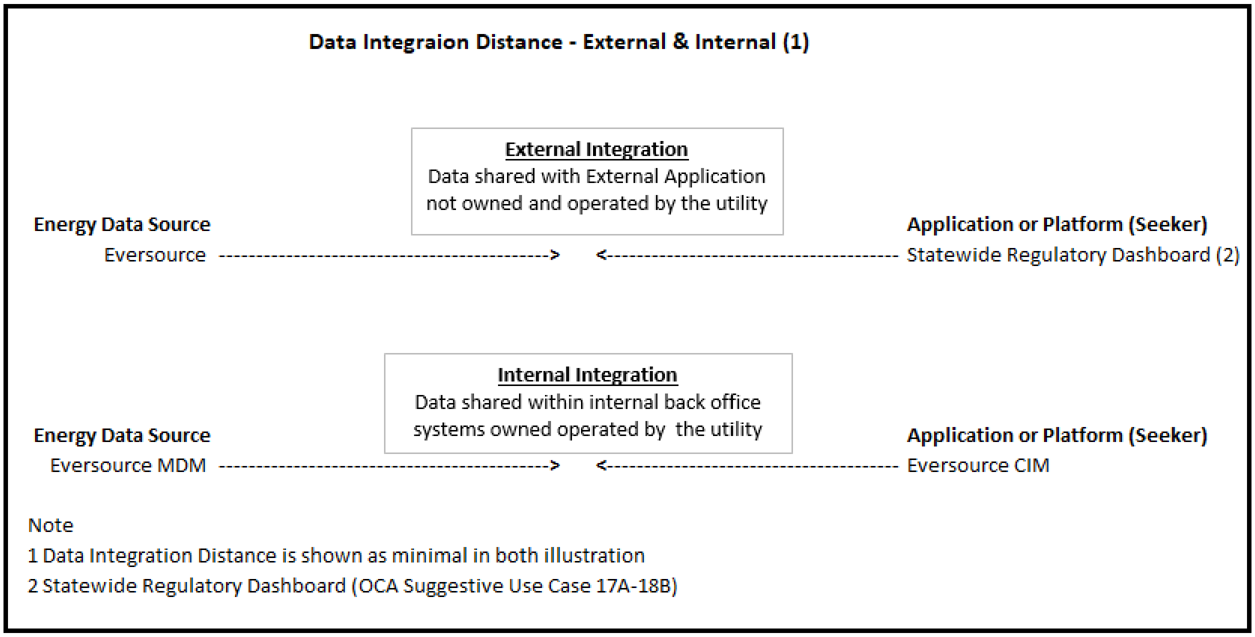
Q. Are there different types of data sharing integration scenarios?

A. For my testimony, I am dividing integration into two distinct categories: external and internal. It is an important concept because my testimony and recommendation is focused on external integration.

Q. Explain the two types of integration, external vs internal.

A. I am providing Figure 5, "Data Integration Distance – External & Internal," to illustrate the two types of integration. The upper part of the diagram is "External Integration," the scenario we are focusing on in this docket where data is being shared among different organizations' Applications. Though not yet possible, in the Figure's example, Energy Data in Eversource's datacenter is being uploaded to a centralized statewide regulatory dashboard that exists outside of (external to) Eversource's data center and control. The lower part of diagram is "Internal Integration," illustrating data being shared between two Applications both within the domain of one organization, in this case Eversource's data center. You can think of internally integrated applications as "pre-integrated" systems. For example Microsoft Office, referred to as a suite, is a bundle of Applications in which data flows more freely between each Application. Similar suits are emerging for utilities. I am not an expert in the IT software of pre-integrated Applications. However, it is my opinion and expectation that large pre-integrated suites of utility back office Applications are becoming increasing available. Notwithstanding vendor lock-in, internally integrated back office Applications should improve data integrations *within* the utility. Figure 5 crudely illustrates a conceptual pre-integrated scenario where Eversource's Meter Data Management System (MDM) seamlessly shares data with their Customer Information System. Assuming this is occurring, then there is zero or low Data Integration distance between the MDM and the CIS system (and thus the arrows illustrating this in Figure 5 are converging).

Figure 5: Data Integration Distance – External & Internal



Q. What is the scope of your testimony and recommendations relative to external versus internal integration?

A. The focus of my testimony is external data sharing, which is the sharing of data among separate systems and organizations as shown in the top part of the Figure and labeled "External Integration."[39] The utility of the future, and the emerging digitization of New Hampshire's industry and energy economy, largely depend on the existence and development of external integration as well as the absence of data integration distance between unrelated Data Sources and Data Seekers. See the Vision section, for my recommendation of an external integration solution using a dedicated new Data Platform that provides a common base of Energy Data.

[39] This can also be referred to as separate domains or separate boundaries.

Q. How can Data Integration Distance, and the associated costs, be avoided?

A. While complex and multi-faceted, the presence of three technical factors can help avoid Data Integration Distance. The three technical factors that must exist are: (1) the use of a Common Logical Data Model that organizes and standardizes Energy Data predictably, coupled with (2) the existence of a published API that provides Applications automated access[40] to the data subject to certain privacy and access standards, and (3) a common base that collects the Energy Data from Data Sources based on the Common Data Model.

Q. Does Data Integration Distance exist in New Hampshire?

A. Yes, there is a wide spread systemic gap. The gap represents the various efforts discussed earlier including: manual effort to transcribe data from paper or electronic bills, screen scraping data from utility webpages using customer password and login credentials, proliferation of ad-hoc databases used by stakeholders and energy managers to organize data, consulting and project teams managing data requests to the utility with six month durations[41] resulting in stale data virtually as soon as the data is usable, failed or delayed projects due to lack of interoperable data, data shared in Excel, PDF and other non-standardized formats, and the work and unpredictability associated with building custom non-standardized unsupported interfaces.[42] These examples of

[40] Within certain privacy and access standards (Data Access Framework (DAF) and Data Privacy Framework (DPF)).

[41] Six month waiting periods for a utility to complete a data request were mentioned in testimony at SB284 legislative hearings and during the DE-19-197 prehearing conference and following technical session.

[42] Here is a real world example to illustrate this challenge. During my New Hampshire outreach I met with an energy manager overseeing multiple buildings. An IT request was fulfilled a few years back. The IOU agreed and developed an IT interface between their IOU's data systems and the Database used internally within the Energy Manager's organization (Data Seeker). The ad hoc interface was "unsupported," meaning it was not included in the utility's (the Data Source's) SCLC change management process. As a result, even though the interface initially worked (and successfully resulted in avoided

manual effort and ad hoc work arounds all represent Data Integration Distance.

Q. Are costs incurred if friction in the energy economy and large Data Integration Distance exist in New Hampshire?

A. Yes, there are significant and recurring costs, many of which are hidden in years of status quo data non-interoperability between Data Sources and Data Seekers.

Q. What are the costs of Data Integration Distance?

A. The costs to New Hampshire of Data Integration Distance fall into four categories: (1) custom integration costs, (2) systemic inefficiencies of operations and processes, (3) opportunity costs, and (4) regulatory inefficiencies.

1. Custom integration costs can be either direct or indirect costs. They are the costs spent by manually overcoming the Data Integration Distance (the gap between the arrows on the lower level of Figure 5 “Data Integration Distance – concept”) in order to get the data from the Source to the Seeker. These costs include labor (by software programmers, excel wizards, systems analysts, business analysts, managers), manipulating data at the last minute to force an integration, writing ad-hoc requirements for the utility, phone calls, e-mails, data entry (typing data from paper bills into databases), hired vendors performing screen scraping, cutting, pasting, and other manual data manipulations. Once data is received, there are often costs in either modifying an Application to interface with the data or modifying the data to interface with the Application. This process then has to be replicated by both the Seekers and the Sources on

manual keying of data from paper statements) it stopped working suddenly when an unrelated change was made to the IOU’s back office systems. The manager ultimately returning to the old process involving his staff keying in data contained in paper statements. A project to correct the interface was not initiated.

regular time intervals. In all, a small army of people spend time overcoming Data Integration Distance. Indirect costs include added overhead of computers, local databases, database administrators to manage databases, systems administrators managing servers, datacenter infrastructure and space for servers hosting, software licenses, and printing costs (when bills are printed for data entry).

2. Systemic inefficiencies of operations and processes occur because they have been created around the minimal low value data that emerges from the current outdated data sharing systems.[43] Data Seekers are forced to perform business and technology tasks without having the data needed to do the task efficiently.

3. Opportunity costs occur in a variety of ways including: investments that are not made because the cost of obtaining the requisite data from the utilities is too high, project cancellations due to the delays in getting data, and the current road blocks to creating innovative Applications from a lack of data availability.[44] Opportunity costs for the SB284 data platform should also include the years of lost opportunity if the market is unable to use the rich new sources of data that will be generated from grid modernization investments as anticipated in IR 15-296, because a data collection system (the Common Data Model, published API, and common base of Energy Data in the SB284 Data Platform) was not built to capture and share data. There is a cost to moving

[43] Assuming the proposed Data Platform is built, it is relevant, from a value perspective, to understand the impact data will have on many existing outdated manual process throughout New Hampshire's energy economy that have been created over the years (a form of the hidden friction already discussed). Part of the hidden value of a Data Platform is the ability to reduce entrenched hidden inefficiencies by either replacing them with efficient processes, or perhaps eliminating the need for the process to exist at all (full automations).

[44] Innovative Applications, products and services, including those in the 55+ OCA Suggested Use Cases, can be visualized to exist in Figure 11 "Proposed Build" in the upper area "External Applications & Platforms." These Applications are largely on-hold due to lack of data access and represent an opportunity cost to New Hampshire.

forward with expensive grid modernization rate based infrastructure investments and ratepayer funded energy efficiency programs that do not include the ability to collect, analyze and share the data created by these investments effectively. IOUs exist to maximize returns to shareholders and are therefore structured first and foremost in a manner calculated to achieve that objective. Having those same IOUs be the only source of data for the competition that threatens IOU profitability is a reliance that is inherently limiting. It is simply unreasonable to expect IOUs to act against self-interest in such a fashion, no matter how earnestly they might proclaim a willingness to do so.

4. There is a cost to the regulatory inefficiencies resulting from a lack of data transparency. Lack of access to data is a roadblock that prevents streamlining and automating various data aspects of energy-related proceedings and regulatory processes. Data sharing for regulatory purposes could add efficiencies from robust and interactive data access during technical sessions, instantly verifiable information for discovery purposes, automation of rules compliance, verification of such compliance, and real-time data access to PUC Staff preparing recommendations or agendas. This could substantially decrease the costs to utilities (and therefore ratepayers) for a variety of regulatory requirements.

Q. Please summarize the current status of data sharing in New Hampshire's.

A. New Hampshire has significant Standardized Integrated Data Sharing Distance and as a result significant time and costs are incurred attempting to access data. In addition, there are significant opportunity costs from lost investments. Avoiding these costs is possible using data sharing technologies that already exist and are widely used in many other industries. New Hampshire does not have a data sharing platform. Currently, data sharing is performed using legacy utility IT systems that are not designed to interoperate statewide. Modern Applications (the hundreds of Use Cases filed by parties in this docket) cannot be developed due to a lack of data access. Because of this, from a technical level, New Hampshire is essentially disconnected from using data in a way that can transform industry and our economy. New Hampshire needs a dedicated statewide data sharing platform to meet current and future needs.

At the same time, passage of SB284 and the litigation of this docket places New Hampshire in a strategically advantageous position, with a green field opportunity. At this point in history, as New Hampshire considers the best direction and strategy to pursue in data sharing, it is important to note that we are not saddled with an existing underperforming data sharing platform with sunk costs. Selecting a prudent but meaningful path to data sharing, as I propose in the Vision section, will not require us to render an existing platform no longer used and useful. New Hampshire has a blank slate and a free hand as it moves forward to create a robust data sharing environment to meet needs of our modern grid as it emerges. This is a good time to think strategically about how this docket will positively influence and enable better results from many other dockets before the Commission.

F. Vision and Public Good

Q. What is your long term vision/goal for data sharing in New Hampshire?

A. My vision is for the creation of a dedicated centralized independent Data Platform that becomes the centralized data access point used and trusted by energy stakeholders and third parties who create Applications (energy service providers and products) that digitally transform New Hampshire's economy. A centralized data access point enables market place innovation and beneficial digital solutions for grid modernization, energy efficiency, distributed energy resources, electric vehicle resources, and new markets and transactions.

Q. Will all the Infiormation Technology required for this vision reside inside the Data Platform?

A. No. Part of the vision requires the IT of the Data Platform and part of the vision requires the IT of the SB284 stakeholders outside of the Data Platform.

Q. What is the economic benefit of access to Energy Data?

A. The economic benefit of frictionless Energy Data (discussed next) is combining benefits that can arise from innovative energy products and services that can exist once access to Energy Data exists. Access to Energy Data[45] can kick start energy and innovations using Applications that could digitally transform the economy as envisioned in grid modernization, sustainable clean energy, and energy-related IOT.

Q. Why do you specifically propose a Data Platform be chosen as the dedicated centralized data access point in your vision?

A. A centralized dedicated Data Platform, rather than a decentralized virtual model, would be the quickest, most flexible, least cost, acceptable risk path towards reducing New Hampshire's Data Integration Distance.

Q. How can a dedicated independent Data Platform provide the benefits of flexibility, least cost that you just mentioned?

[45] "Access" to Energy Data is assumed to be restricted and fall under regulated policy resulting from Commission-approved Data Access Framework and Data Privacy Framework as defined earlier and included in my Recommendation section.

A. The Data Platform will be used to implement an integration strategy known as External Integration, as a least cost acceptable risk strategy for bringing robust data sharing to New Hampshire.

G. External Integration Strategy - Least Cost

Q. You've used the term "least cost" within the title of this section, what costs are you referring to?

A. I am referring to the two sets of IT costs that need to be incurred in order for New Hampshire to realize that full data sharing vision. Therefore in this section, I am referring to the traditional costs of building, testing, operating, maintaining, and upgrading the technology in these two areas shown, (1) IT of Data Platform, and (2) IT of Stakeholder Applications.

Q. Do the hidden costs of the invisible friction described earlier in your testimony fall into the two areas of cost listed above?

A. No. The hidden costs of friction are the systemic costs of inefficiencies associated with poor data access. These hidden costs, made visible in the Data Integration Distance metric, are the cost New Hampshire can begin to avoid if Standardized Integrated Data Sharing is realized using an External Integration strategy that is implemented through the creation of a Data Platform.

Q. What is External Integration?

A. External Integration is a well-known widely used strategy for integrating data contained in one or more independent systems by utilizing an outside system (an external system). The strategy places the burden of data integration on an independent external system or platform (in this case SB284 Data Platform) while allowing original Data Source systems, i.e. utility and non-utility back office systems and data contained in those systems, to remain in place, largely unchanged.[46]

Q. Is External Integration used today in the electric distribution industry and recognized as a valid strategy by grid domain experts for integration projects?

A. Yes. External Integration is widely used in the technology world, including electric distribution utilities. EPRI states "external integration is still needed and the use of a common integration platform still has many benefits." And they state "common integration technologies and platforms used today have names like Service Oriented Architecture (SOA), Enterprise Application Integration (EAI), Enterprise Service Bus, Web Services."[47]

Q. Why is External Integration strategy a potentially good fit as a solution to New Hampshire's data integration problem?

A. External Integration is an effective strategy that would cost effectively address four technology scenarios that exists across New Hampshire's Data Sources that to varying degrees create significant Data Integration Distance and therefore prevent them from providing Standardized Integrated Data Sharing needed to support Use Cases such as those submitted by parties to this docket. The four challenging technology areas that an External Integration solution can address are:

1. Scenarios where Data Sources lack an adequate interface, possibly due to a lack of

[46] The integration process may include connecting to and exchanging data with existing utility (or nonutility) systems for operational purposes such as updating data in the SB284 data platform. Depending on implementation decisions, on a utility-by-utility basis, this integration process may include the Installation of agreed software and or hardware component(s) to standardize, automate, streamline, coordinate, and test the integrated data sharing process.

[47] *See* Appendix 2 for OCA's Master Use Case "SB284 as a Data Platform," which suggests the Data Platform design be a Service Oriented Architecture.

need for an interface, legacy design, security requirements of the system, management policy, etc. As a result, External Applications are closed out and unable to access data stored inside the Data Source.

2. Scenarios where a Data Source has its data defined by a Data Model incompatible with the Data Seeker's system. The Data Source has the necessary Data Types, but the data is organized and defined in a way that is unusable to the Data Seeker's Applications.

3. Scenarios requiring data records from multiple Data Source systems to be consolidated. Even if Data Models are compatible, it may not be feasible or prudent to consolidate data from multiple Data Sources internally.

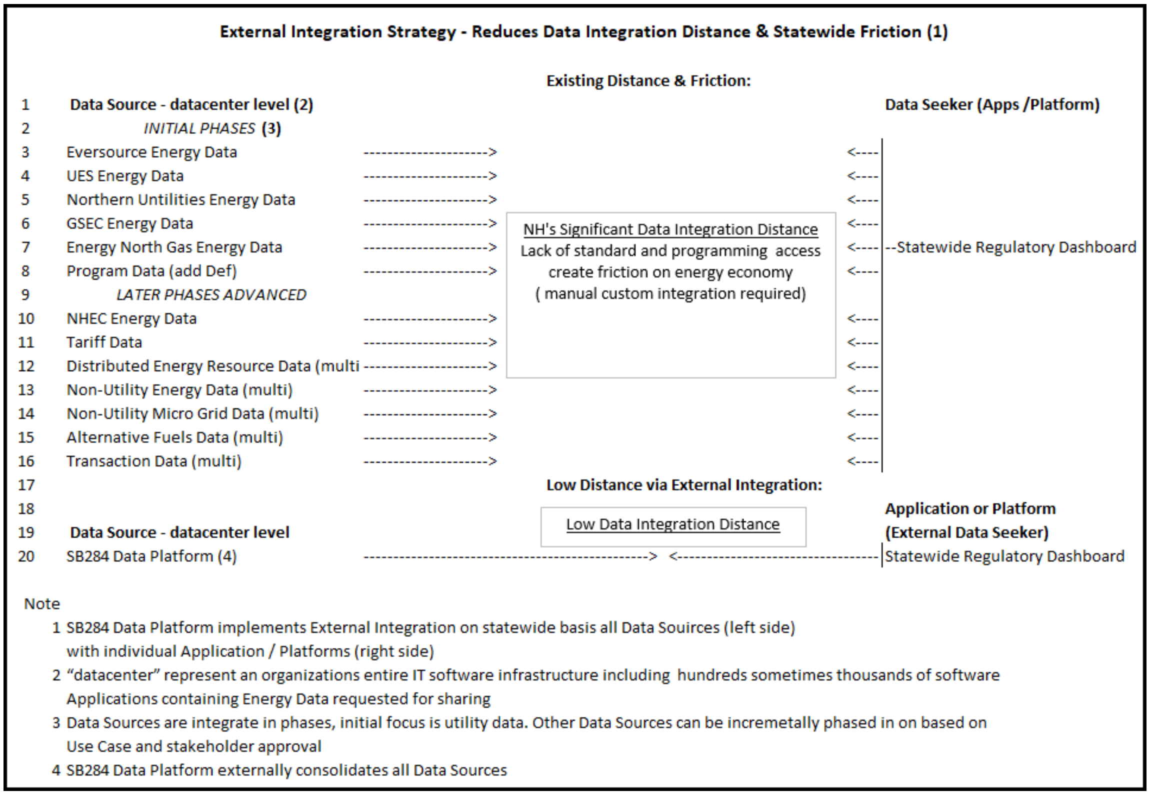
4. Scenarios where different Data Types (and Data Entities) are stored at multiple Data Source systems and need to be consolidated into one Data Model.

Using a common strategy, an External Integration strategy, which addresses the multiple technology problem scenarios I have listed above, will lead to a statewide data sharing solution.

Q. What does "statewide data sharing solution" mean relative to Data Integration Distance?

A. It means the Data Integration Distance is reduced on a statewide level for all Data Sources and Data Seekers. Figure 6, "External Integration Strategy – Reduced Data Integration Distance & Statewide Friction," illustrates the problem (at the top) and the solution (at the bottom).

Figure 6: External Integration Strategy - Reduced Data Integration Distance & Statewide Friction



At the top left, in rows 3 to 16, are a multitude of diverse Data Sources. At the top right is a Data Seeker such as a proposed Regulatory Dashboard. Note that any of the 55+ OCA Suggested Use Cases in Appendix 2 could be substituted here. In this example, for the Dashboard to get its data, it must separately integrate with each of the Data Sources on the left. However, due to significance distance between the dashboard and the individual sources, the task become large and complex and too costly, and will likely be deferred to another time, which will, in the process, create hidden opportunity costs.[48]

[48] While the deferred Regulatory Dashboard projects avoids immediate costs (project team, software development, etc.) an invisible opportunity cost is created – elimination of efficiencies, savings and benefits that would occur if the Dashboard

The statewide solution is shown in the lower area at row 20. The Data Platform becomes a single data access point for the Regulatory Dashboard, or many other SB284 Use Cases. Because there is zero or low Data Integration Distance, as shown in row 20, the dashboard project would be significantly less costly to build and operate, due to the data Platform's existence, and therefore becomes a potentially viable project.[49] As the Data Platform matures in future years, additional Data Sources may become available (for example rows 11-16 on the left). More data may result in the Regulatory Dashboard becoming increasingly beneficial.

Q. What two factors make External Integration a least cost strategy?

A. External Integration entails physically creating a Data Platform in a new environment where the two problems defined earlier (poor data and inadequate access) are solved. Cost advantages are possible if the solution space is moved away from a fragmented decentralized set of complex IT environments, to a new centralized external solution space dedicated to solving New Hampshire's data sharing problem as discussed in Section II.D. "New Hampshire Status and Problem Statement." The cost advantages of the External Integration strategy would be significant and would stem from two factors: (1) greater technical feasibility, and (2) economy of scale. These two factors will drive savings when building the Data Platform and when building the many External Applications that will connect to the Data Platform.

became an approved project and was created. Opportunity costs are discussed as one of the four costs associated with friction and Data Integration Distance.

[49] Building the Regulatory Dashboard using the External Integration strategy (Data Platform) is one example of least cost of "IT of Stakeholder Applications" discussed at the beginning of this section.

Q. Briefly discuss these two factors, technical feasibility and economies of scale, that result in External Integration strategy being a least cost strategy.

A. I will briefly discuss the factors in four parts while referring to Figure 6:

The first factor is technical feasibility. For all practical purposes, and assuming the goals is Standardized Integrated Data Sharing, it would be impossible to individually integrate each of the Data Sources illustrated on the left side of in rows 1-16 with increasing numbers of External Application (Data Seekers) on the right. The hundreds or thousands of different database and data model, different interfaces, different organization, plus the overall lack of APIs create an enormous problem to solve within that same disparate environment. Advantages are gained by moving the solution space to a clean external environment where existing well known technologies are fully capable of addressing the problem and are widely used in major industries worldwide.

The second factor is economies of scale of a centralized platform. Building and operating one centralized data sharing platform is least cost when compared to building and operating multiple decentralized data sharing platforms. The cost of designing, building, operating, testing (including User Acceptance Testing) upgrading, and regulating a centralized Data Platform will be significant. Repeating these same costs by having each utility and non-utility Data Source, illustrated in rows 1-16 on the left, create and operate separate platforms is more costly due to duplicative effort and infrastructure.

The third factor is economies of scale for Stakeholder Applications. It will be less costly for each of the stakeholder External IT Applications, to access data from one centralized Data Platform. Requiring stakeholders to design, build, and operate their Stakeholder Applications that must integrate with multiple decentralized data sharing platforms (multiple Data Sources) is significantly more costly. Whereas, both the time and cost to Stakeholders building Applications that only need to integrate to one access point (one Data Source) is significantly lower. Therefore, the existence of a centralized access point capable of generating statewide datasets, for utility and non-utility data sources, creates economies of scale for Stakeholder building and maintained External Applications that require Energy Data.

The fourth factor is the benefit of avoided risk. There is a potentially lower risk to back office utility IT systems when data sharing occurs externally, in an environment physically and logically separated from the utility systems. If potential risk to on-going mission critical utility IT systems can be reduced using an External Integration model, i.e. major portion of IT functionality is located outside of the utility legacy mission critical IT systems, then there may be lower project risk and potential cost savings.

Q. Based on your experience, what is required of this "solution space" for the strategy to succeed?

A. The three things required for a project to succeed are: (1) proper people (knowledge and skill sets), (2) proper environment (technology and infrastructure), and (3) proper business framework.

Q. How will External Integration work using the proposed Data Platform?

A. The SB284 Data Platform, as discussed in the Data Platform section, will use a Common Data Model and well know technologies that can implement a least cost External Integration strategy.

H. Data Sharing Vision Summary

Q. Please summarize your data sharing vision.

A. The data sharing vision for this docket is for a dedicated independent Data Platform that becomes the centralized data access point used and trusted by energy stakeholders and third parties who create Applications and Platforms (energy services and products) that digitally transform New Hampshire's economy. In order to be a trusted access point useful to innovative beneficial Stakeholder Applications, we need a strategy to create Standardized Integrated Data Sharing in New Hampshire. The least cost strategy for providing such a robust level of data access is known as External Integration which makes utility data interoperable, while limiting risk to current IT systems used by the utilities to run their business. This External Integration Strategy will be implemented by the creation of the proposed SB284 Data Platform discussed in the next section.

III. Data Platform

A. Introduction and Organization

Q. How is your testimony on the Data Platform organized?

A. In this section, I start by explaining how certain key concepts are interrelated and form the basis of the Data Platform. I then discuss Use Cases and how data analysis must be part of the Use Case process. Analysis of the data needs of each use case is a key to a successful platform, while failure to analyze the data may lead to a Data Platform that underperforms or fails due to poor data quality. I then describe in deeper detail each of those concepts including the difference between physical and logical data, the challenges of physical data, data entities, the importance of a logical data model, and the relevance of NEEDS as an existing model. This is followed by a description of the design principles, prerequisites and steps to build the proposed SB284 Data Platform.

Q. What is a Data Platform?

A. The Data Platform is the physical asset comprised of hardware and software that will become New Hampshire's centralized data access point, used and trusted by energy stakeholders and third parties. The Data Platform is the physical implementation of the External Integration strategy discussed earlier. The implementation of the External Integration strategy is the most cost effective way to eliminate the Data Integration Distance that is causing such friction and inefficiencies in the New Hampshire energy economy.

Q. What are the key functions of a Data Platform?

A. As illustrated in Figure 1, the two key functions of the Data Platform are (1) to load data and (2) to share data. Loading data from multiple Data Sources and sharing that data with different Data Seekers requires Standardized Integrated Data Sharing.

Q. What are the key concepts that should be understood before looking at details of the Data Platform?

A. There are three areas that need to be described in greater depth before diving into the Data Platform details. The first is how Use Cases help inform what the Data Platform becomes. The second is the distinction between physical data and logical data. The third is the challenge of physical data. The fourth area is the role of Data Entities, the Logical Data Model and the NEEDS model.

B. Use Cases – Follow the Data

Q. How do you suggest approaching the Use Cases developed by the OCA and other stakeholders in this docket?

A. I propose a "follow the data" strategy. Specifically, I recommend thinking of Use Cases as Applications all performing the same three actions which are (1) receive data, (2) do something with the data, and (3) send data. Regardless of the particular Use Case (a Data Platform, a Reporting Portal, a regulatory dashboard, a DER Registration Platform, an ISO-NE Settlement System, a CCA system or sub-system, or a Green Button Platform operated by a national vendor), these Applications are all in the business of getting data, manipulating data, and sending data. In considering the Use Cases that become part of the record in this docket, I recommend focusing on these actions – particularly the first action, "getting data."

Q. What are Use Cases?

A. For purposes of my testimony Use Cases, can be split into two categories: (1) stakeholder Application Use Cases (see Appendix 2, "OCA Suggested Use Cases with Brief Description") and (2) Platform Use Cases which is the OCA's Master Use Case in Appendix 2.

Q. What is a stakeholder Application Use Case?

A. A stakeholder Application Use Case describes a beneficial External Application that the SB284 Data Platform will support by providing it data. OCA Suggested Use Cases with Brief Description in Appendix 2 are examples of stakeholder Application Use Cases. These stakeholder Applications are also illustrated in Box 3C of Figure 1 and shown from a system design perspective in Figure 11 "Proposed Build" in the layer labeled "External Applications & Platforms – proposed Phase 1."

Q. What is a platform use case?

A. For purposes of my testimony, a platform Use Case describes a Data Platform that holds organized data that will be requested by supported stakeholder Applications. The OCA Master Use Case "SB284 as a Platform" in Appendix 2 is a platform use case and is illustrated in Figure 1, Box 3A, and shown from a system design perspective in Figure 11 "Proposed Build" in the layer labeled "SB284 Statewide Multi-Use Online Data Platform.

Q. What happens if we don't analyze the data needs of stakeholder Application Use Cases, particularly their "get data" action?

A. A major risk that can lead to a Data Platform not being able to become a trusted source of data is poor data quality. Analysis of potential stakeholder Applications informs the design of the Logical Data Model, discussed shortly in my testimony. Failure to design a robust Data Model is a risk that is difficult to mitigate or recover from. The strategy to "follow the data" and analyze that data is critical. The process of analyzing the data needs of Stakeholder Applications and designing the Data Platform to conform to those data needs is also critical.

Figure 7: Full Data Sharing Vision – Collaboration of Use Cases in Different Boundaries

Full Data Sharing Vision - Collaboration of Use Case in Different Boundaries (1)

Part of vision statement derived from OCA's Master Use Case"	Part of vision statement derived from Stakeholder Use Cases
"centralized data access point used and trusted..."	*"...by energy stakeholders and 3rd parties who create Applications and Platforms (energy services and products) that digitally transform NH's economy"*

BOUNDARY A **OCA's Master Use Case**			**BOUNDARY B** **Stakeholder Use Cases (2)**
#1 "SB284 as a Centralized Platform" (Data Source)	-------------->	<----------	#1 Energy Efficency Application (Data Seeker)
#1 "SB284 as a Centralized Platform" (Data Source)	-------------->	<----------	#3 Green Button Platform (Data Seeker)
#1 "SB284 as a Centralized Platform" (Data Source)	-------------->	<----------	#4 Community Energy Dashboard (Data Seeker)
#1 "SB284 as a Centralized Platform" (Data Source)	-------------->	<----------	... ...more (Data Seeker)
#1 "SB284 as a Centralized Platform" (Data Source)	-------------->	<----------	#54 Asset Value - Tracking Tool (Data Seeker)

Note

1 Full Data Sharing Vision: "centralized data access point used and trusted by energy stakeholders and 3rd parties who create Applications and Platforms (energy servicers and products) that digitally transform NH's economy" bates xxxx

2 Stakeholder Use Cases shown are illustrative and were developed by OCA based research and outreach Ref SB284 2 page brief.

C. Logical data versus physical data

Q. Why is it important to understand the distinction between "logical" and "physical" in this docket?

A. Understanding the distinction is critical to grasping important concepts of data sharing, including the term "logical data model" explicitly referenced in RSA 378:51 I(d). The key is to understand that data can be viewed from two perspectives: physical and logical.

Q. What is the main distinction between logical and physical data?

A. Logical data is an abstract representation of objects, with no actual connection to the physical world. Logical in this context means not physical, where physical is the real world. Logical data is often represented by Data Types or Data Categories. For example, "customer," "transaction," "market," "DER," "system," and "policy" are all Data Types or Data Categories that represent logical data. Physical data has physical aspects. For example, Data Sources with databases holding database records. Physical data is data physically entered (hand typed, imported, linked) into rows of database tables where each row is called a record.

Q. Discuss the concept of physical data.

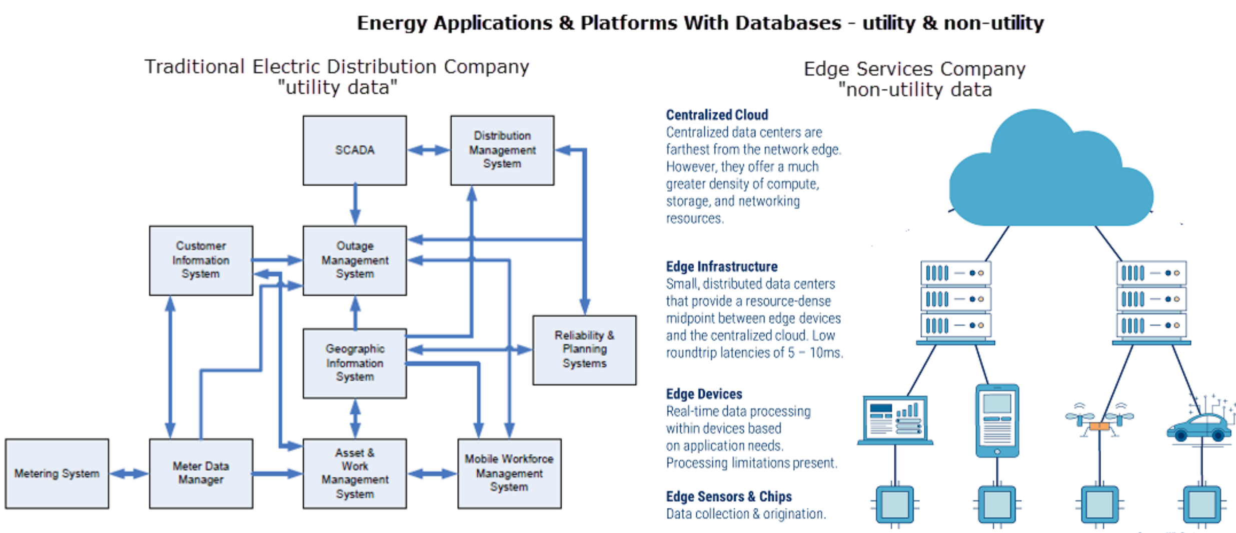
A. Physical data comes from Data Sources. Data Sources can be bifurcated into utility and non-utility sources. Often these sources can provide similar kinds of data. So the same kind of data could be utility data or non-utility data depending on the Data Source.

The physical data or records from different utility and non-utility Data Sources may exist inside hundreds or thousands of different databases and applications (software) installed on different servers (hardware) or cloud locations based in different data centers. Databases often have different types and manufacturers such as Microsoft, Oracle, IBM, open-source, etc. Databases organize data in many different ways based on each database's physical data definition.

Q. Provide an illustration of the different ways physical data is held by utility and non-utility sources.

A. Figure 8 shows a conceptual illustration of where databases holding data from utility and non-utility sources may be located. The left side represents a distribution utility with traditional system such as meter data management system, customer information system, or geographic information system. Each of these systems have databases and are storing data according to their individual and sometimes often unique or proprietary Data Models. The right side is a very conceptual illustration of non-utility data. For purposes of my testimony it represents data from an energy service company, potentially serving Use Cases.

Figure 8: Energy Applications & Platforms with Databases – utility and non-utility



Q. What are the key challenges to using existing physical data?

A. The challenge to physical data is that it is located in different databases. Databases storing data can have different Physical Data Models, which leads to data interoperability issues and data silos as well as software stovepipes when there is a need to share data between Applications (for example sharing data in a utility CIS with a CIS in a CCA Platform. There are likely thousands of databases and Applications physically storing Energy Data in New Hampshire utilities and non-utility Data Sources. Each of New Hampshire's IOUs can have unique IT systems, purchased from different vendors, different datacenters operating on different operating systems, different Applications, Data Models, and processes. This uniqueness results in a lack of consistency and is shown on the left side of Figure 8. As Figure 8 also shows, non-utility energy related organizations also host databases storing New Hampshire Energy Data in their own unique data environments. A Logical Data Model can be a solution when there is a need to consolidate data from so many different formats.

Q. You mentioned Data Entities earlier. How are Data Entities related to the challenge of physical data?

A. Data Entities are the data physically contained in the database that represents entities such as customers, locations, measurements, etc. The challenge is that if, for example, two utilities or non-utilities both have a customer entity in a Database, and each database has a different Physical Data Model (aka the structure that dictates how that Data Entity is defined) you now have two unique Data Entities that represent the same thing. Another example that occurs regularly today is two utilities have the same customer, but they enter the address in slightly different ways, and now that customer entity cannot be matched by address.

Q. What is the impact of multiple and inconsistent definitions of similar Data Entities?

A. Inconsistent Data Models resulting in duplicative but inconsistent Data Entities. This is one of the two causes of Data Integration Distance preventing robust data sharing in New Hampshire.

Q. How do we address this issue of inconsistency across different Data Sources?

A. Currently Data Seekers have to address this inconsistency manually by processing the information they receive by excel or .pdf into their own database. So whether a Data Seeker is a customer, a consultant providing Evaluation, Measurement and Verification (EM&V) reports for energy efficiency program administrators, or a competitive energy service provider, a Data Seeker put tremendous resources in to trying to merge data from different Data Sources. Alternatively, a Data Platform can be created that brings in the data from different sources and makes it available to Data Seekers. To be successful the creation of that centralized statewide database, the Data Platform has to be based on a logical data model that can overcome the inconsistencies and silos.

Q. Help us understand what the terms 'logical data' and 'logical data model' mean.

A. It is useful, and often insightful, to ignore the physical constraints of data in terms of who, where and why the data is physically held, in order to look at data more abstractly and strategically. For example, one may want to look at kilowatt-hour measurement across all sources statewide, or segment them into logical groups such as kilowatt-hours for buildings with installed energy efficiency measures, or buildings that have DER of type storage, or buildings not on a time-of-use rate but equipped with a smart meter, or a combination of criteria.. A similar logical example is to count the DER records across both the utility and the non-utility data sources, including asset nameplate attributes, in order to calculate the percent of DERs that exists behind the meter in New Hampshire. The point is that logical data can be used to gather data across different Data Types or Data Categories that can be related logically to make possible additional information, calculations and insights. This type insightful data analysis is not readily available using legacy utility IT systems designed for different purposes.

For these reasons, to gain transparency and insight, the concept of logical data is fundamental to the development of a Logical Data Model as required for the Data Platform. Next I will discuss Logical Data Models and a proposed new Logical Data Model called NEEDS which I recommend form the basis for defining a statewide centralized database that becomes a "common base of energy data" for New Hampshire. The SB284 Data Platform should be designed to receive data and share data that empowers consumers and enables the creation of new and competitive energy products and services.

D. Logical Data Models and the Role of NEEDS

Q. What is a logical data model?

A. A statewide Logical Data Model is a requirement of the Data Platform under RSA 378:51(d). As referenced in the DE 16-384 settlement agreement, the Logical Data Model is the documentation of tables and relationships within the energy database.[50] In my testimony a Logical Data Model is another name for a Data Model. "Data Model" or "Logical Data Model" is an abstract representation of (1) a selected set of real world objects called Data Entities (such as "Customers," "Accounts," "Sensors," or "Owners"); (2) their attributes (account number, meter ID); (3) their associations and relationships to each other (e.g., meters have measurements); and (4) data constraints (e.g., measurements must have duration). A Logical Data Model describes a physical world set of objects without concern to where these object are physically stored such as in a physical database used by an Application.

[50] *See* Appendix 1 at 10 for the language in Section 7.7 of the Settlement Agreement.

A Data Model is not a physical database, but can be used to define such a database. Logical Data Models can describe formats for data packets transmitted between two separate interoperable IT systems. Essentially, the Logical Data Model is the agreed definition for the data and data relationships. As such, it is the key to enabling current and future Use Cases. The Logical Data Model should be documented and published so that it can be used by multiple parties.

Q. Has work on a Logical Data Model been done in New Hampshire?

A. Yes. Under the Settlement Agreement in DE 16-384, a Data Working Group was established and worked on development of an initial beta version of a Logical Data Model. This work can be leveraged for the development of a Logical Data Model for the SB284 Platform. The result would be the NEEDS (New Hampshire Electric and Gas Energy Data Standard) Model.

Q. What is NEEDS?

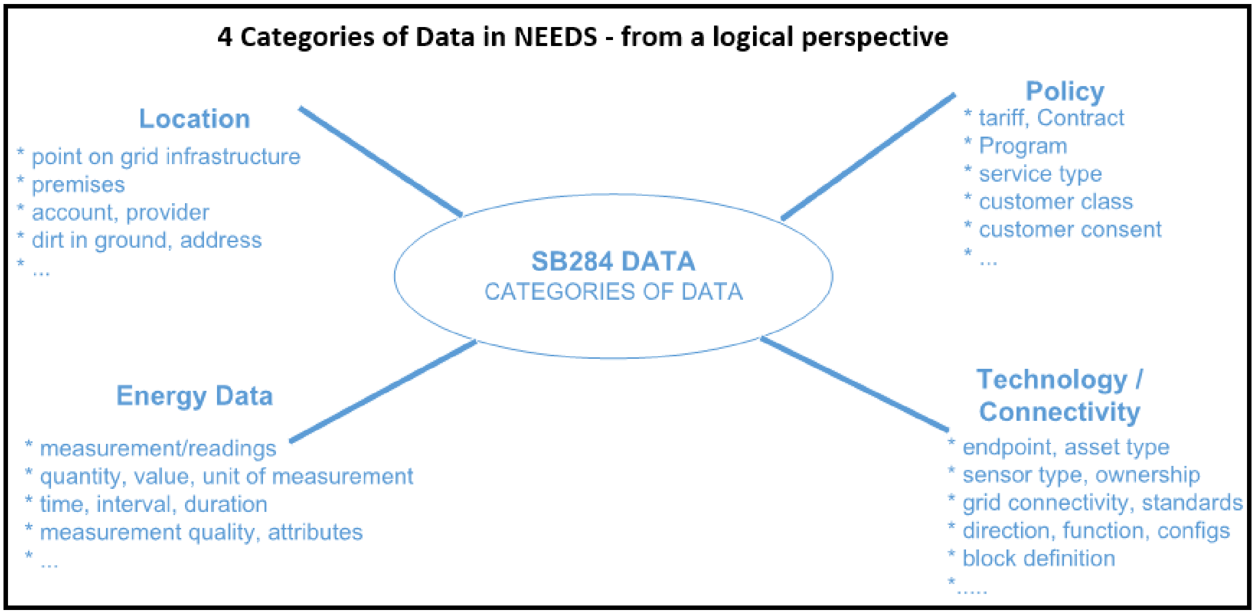
A. NEEDS is a way to refer to the Logical Data Model to be developed for the SB284 Data Platform and defines the associated database and physical data model. NEEDS will help define the scope of the Data Platform's two key functions to load data from Data Sources and share data with Data Seekers.[51] The Logical Data Model and associated Data Entities should be designed to enable Standardized Integrated Data Sharing for Use Case Applications.

NEEDS will contain at least 35 Data Entities including those listed on page 12 of Appendix 1. In more general terms the Data Entities can be categorized in four ways: (1)

[51] Analysis of data that is needed from Data Sources for Data Seekers is determined in Use Case analysis.

location, (2) policy, (3) technology/connectivity/configuration, and (4) Energy Data (granular, with time and location dimensions). Figure 9 "Categories of Data in NEEDS "reflects a logical way to view Data Entities in NEEDS based on how they are related with each other.

Figure 9: Four Categories of Data in NEEDS – from a logical perspective



Q. How are Data Entities treated differently in the proposed SB284 Platform than existing utility and non-utility databases?

A. NEEDS will relate Data Entities differently than how they are related in the databases of the utility and non-utility sources. In Figure 9, the four lines extending from the center, and all the categories of data listed in the four areas, are related in the SB284 database itself. This allows powerful data sharing where any data from any category can be included in a complex query statement. For example a Data Seeker, such as the PUC, could request a dataset for all customers statewide, by circuit, with a DER of type storage, who are not on a time-of-use tariff, who have an AMI meter, and have a particular load shape during peak hours.

The Data Entities in NEEDS should be modeled specifically to implement the proposed External Integrations Strategy as discussed in the "External Integration Strategy – Least Cost" section, and the needs of supported stakeholder Application Use Cases discussed in the "Use Cases – follow the data" section. In addition, data should be shared at a higher level than exists today in New Hampshire. I have defined this higher level as Standardized Integrated Data Sharing and discussed it further in the "New Hampshire Status and Problem Statement" section.

The Data Entities associated with the hundreds and even thousands of databases of Data Sources and Data Seekers can be different even for the 'same' Data Entity (such as Customer or Meter) because each of those databases associated with the Data Entity has a different physical data model designed to support different purposes and different Use Case Applications such as a meter data manager. The goal of the statewide SB284 Data Platform is to develop NEEDS, a Logical Data Model, to organize Data Entities uploaded from a multitude of different Data Sources and different databases, into a standardized database that can then be queried as described above to generate information required by stakeholder Applications/Use Cases/Data Seekers.

Q. What is the key to the development of NEEDS for the SB284 Data Platform?

A. The key is to create a forward looking architecture that has the ability to grow in order to accommodate future Use Cases as contemplated in the OCA's Use Case suggestions for immediate and future implementation in Appendix 2. NEEDS will determine what data *can* be loaded to the SB284 Data Platform – from a data architecture perspective. Data that is being loaded to the SB284 Data Platform is determined by Use Cases authorized in each phase. If a particular Data Type or Data Entity is not part of the architecture at the time, then that data cannot be loaded to the Data Platform. NEEDS does not determine what physical data *is* being uploaded loaded to the SB284 Data Platform. The actual levels of physical data loaded to the Data Platform are determined by the Use Cases being supported in that phase.

Q. Why have a forward looking architecture and incorporate Data Entities or attributes that are required by the Use Cases?

A. The NEEDS architecture must be informed by national standards, New Hampshire goals, and stakeholder Use Cases to allow phased increases in data sharing to support the anticipated additional data sharing needs discussed in a variety of dockets from grid modernization to electric vehicles, energy efficiency and community choice aggregation.

Phased increases in levels of data sharing are prudent and allow initial projects, often the low hanging fruit, to be kick started immediately. Failure to consider New Hampshire's future data sharing requirements as the system is being designed, and only focusing on phase one needs, increases the future risk of the SB284 Data Platform becoming obsolete, or in need of costly upgrades, or necessitating a less efficient and more costly process of running multiple platforms side by side.

Q. Please provide additional background and insight on how NEEDS should be developed.

A. NEEDS should be designed as a generic set of Data Entities to which all New Hampshire utilities and non-utilities are able to map their data. This will mean that Data Sources with differently defined Data Entities and differently structured databases will be able to upload their data to the SB284 Platform if they have complied with the NEEDS parameters.

The NEEDS design and architecture should be based on: (1) an analysis of well-known models including CIM and ESPI, (2) analysis of New Hampshire specific aspirations including policies, Commission Orders, laws, and energy visions and requirements documented in stakeholder Application Use Cases, (3) analysis of data sharing issues in other states and discussions with national experts, (4) the requirements of RSA 378:50-54, and (5) the potential Use Cases or Applications that will benefit the New Hampshire economy and ratepayers.

Q. Please summarize why the development of NEEDS as a Logical Data Model for the SB284 Platform is so important.

A. Currently, there is an extensive level of non-interoperable New Hampshire Energy Data contained in the hundreds to thousands of databases managed by a variety of largely independent groups of Data Sources. The concept of a centralized Data Platform is to strategically identify a subset of this disparate data and unify it for use by a variety of Data Seekers such as ratepayers, energy service providers, utilities, regulators, and communities.

E. SB284 Data Platform: Design Principles, Prerequisites, Build Steps

Q. Please summarize this section of your testimony on the SB284 Data Platform.

A. In this section I define a Data Platform architecture and operational process that will share data from new environments while limiting risk to existing utility operations and legacy systems.[52] The objective is to facilitate the reception of relevant data from Data Sources such as the regulated utilities and load it to a centralized database (New Hampshire's SB284 Data Platform or "common base of energy data"). As data is uploaded and stored inside the SB284 Data Platform architecture it is merged (Externally Integrated) with data from other sources to deliver

[52] External Integration allows the Data Platform to operate outside the boundaries of the existing utility legacy IT systems as discussed in the "External Integration – Least Cost" section.

the robust data sharing required for New Hampshire to participate in an evolving energy economy as envisioned in the grid modernization docket, IR 15-296. I will discuss the architecture and process next.

Q. Summarize the SB284 Data Platform as you recommend it.

A. The Data Platform will securely collect, organize, protect, and share Energy Data. The Data Platform will be based on a statewide Logical Data Model and will be designed to conform to the privacy and cyber security policies[53] prescribed by the Commission. There are software, services, databases and APIs that make up the Data Platform.

Q. What are the design principles of the Data Platform?

A. There are three basic principles that should be adhered to when building the Data Platform. First, the Data Platform should be built on a versioned data model. This means the database of the SB284 Data Platform will conform to an agreed version of NEEDS and I encourage the Commission to require Data Sources, including the utilities, to adhere to the same version of NEEDS as is used to build SB284 Data Platform. Second, the Data Platform needs a Service Oriented Architecture. Lastly the key to the Data Platform development is the Application Program Interface that is developed for interaction with Data Sources and the API developed for interaction with Data Seekers.

[53] Privacy and cybersecurity policy would be consolidated and documented in the proposed Data Access Framework (DAF) and Data Privacy Framework (DPF) which I introduced in Section I.

Q. What are the key features for the Data Platform?

A. The Data Platform has seven core features allowing it to operate as a Platform and to support multiple Stakeholder Use Cases and External IT Applications and Systems. These features are: (1) granular data and billing data, (2) time-of-use data, (3) multiple unit-of-measurement data, (4) multi-utility data, (5) multi-fuel data, (6) statewide index data including location and external source data, and (7) an external interface.

Q. What are the benefits of the seven core features listed above?

A. These features support multiple and disparate data needs of beneficial IT applications, platforms, processes, workflows and process improvements collectively referred to as "External IT Systems." Examples of beneficial applications are contained in the OCA Suggested Stakeholder Use Cases in Appendix 2. Because the features outlined will create an expandable architecture that can accommodate current and future use cases, there are a higher number of benefits to consider when evaluating the costs of building the SB284 platform.

Q. What are the prerequisites that must be completed prior to building a Data Platform?

A. There are four prerequisites that need to be determined through an established governance body and approved by the commission: business requirements, NEEDS, DAF, and DPF. These prerequisites are policies and documents that essentially codify the goals and policies of the Data Platform.

Q. Briefly discuss the four milestones: Business Requirements, NEEDS, DAF, and DPF.

A. The business requirements should be properly documented and incorporate the Use Cases, data requirements, and the Sequence Diagrams[54] that are recommended for the integration project. SB284 Data Platform is an integration project. It integrates data between Data Sources and Data Seekers.

The NEEDS Common Logical Data Model should be a tested, versioned with technical documentation. Data Models, such as NEEDS, can be managed and administered through an appropriate technical organization (separate from the governance bodies I discuss later in my testimony) that educates users and manages the model. NEEDS should include the seven core features mentioned previously and be organized in a way that supports diverse Stakeholder Application Use Cases. Considering future potential Use Cases will lead to a data architecture that provides flexibility to adapt to evolving data needs of grid modernization, energy efficiency, and new energy service markets.[55]

The Data Access Framework and Data Privacy Framework should be centralized policy documents approved by the Commission, likely by rule. The DAF and DPF need to define comprehensive restrictions on users (Applications) of the Data Platform. They are the mechanism for documenting strict limitations such as exactly who can access data, under what circumstances, such as consent or restrictions, etc.), the level of access, and how the data is secured and protected. The DAF should essentially govern who can see what data, under what conditions. The DPF should guide all privacy rules, consent, cyber security, and customer data

[54] A sequence diagram is a form of UML diagram that shows how processes and actors integrated with each other.

[55] New Hampshire's energy markets represented as "New Hampshire Economy enabled consumers and industry" shown in the top layer of Figure 11 (Proposed Build).

protections.

Q. Why is this documentation critical as the prerequisites to building the Data Platform?

A. IT projects that lack adequate documentation can have higher risk of cost overruns and or project failure. Failure to perform due diligence, such as creating documentation and specifications, can result in a vendor bidding on and building a Data Platform that does not meet needs. Poor documentation can also create difficulty when determining if the Data Platform is functioning properly prior to going live.

The documents are critical to supporting important efforts going forward. The business and technical specification documents will be provided to vendors bidding on or building parts or all of the system. They can be used as attachments to a potential RFP and/or RFI. The documents will provide the guiding instructions to development teams writing the software, services, databases and APIs that make up the Data Platform. There are areas that require User Acceptance Testing (UAT) on the Data Platform prior to accepting delivery and making payment. This process will include approval of test scripts, approval of results (pass / fail), and running the tests based on the agreed test scripts. The documentation can then be used as evidence in any dispute with contractors who build part or all of the systems. In short, the documentation is key to the mitigation of risk.

Q. When the above prerequisites are complete what are the high level steps that must be completed to reach the goal of a functioning Data Platform used by stakeholders and their Applications?

A. The three high level steps are (1) building the Data Platform, (2) loading the Data Platform with Energy Data, and (3) making the Energy Data accessible by query to approved Data Seekers.

Building the Data Platform to support the two key functionalities of loading and sharing data. This includes the hardware and software based on the Logical Data Model and the other requirements established in the prerequisites I just mentioned. The data loading step creates the common base of energy data required by the statute. The last step is for the Data Platform to start sharing data with the agreed list of External IT Applications according to the Standardized Integrated Data Sharing standards. Standardized Integrated Data Sharing will be provided to those stakeholder Applications included in Phase 1.

Q. Briefly discuss Step 1, building the Data Platform.

A. Building the Data Platform is a major topic. I will mention just three of the many important areas to consider in this at this stage: management, testing and phasing.

Traditional IT project management and planning methodology should be used for creating the Data Platform. In general, the systematic and proven process known as the System Development Life Cycle SDLC should be used. The SDLC adds coordination, transparency, and knowledge of status in a way that help reduce risk and improve chances of success. There are six stages to the SDLC that can be directly followed by the project team. The six stages of the SDLC are: (1) requirement analysis, (2) design, (3) development and testing, 4) implementation with change management and versioning, (5) documentation, and (6) evaluation. Many of these stages are identified in RSA 378:51 as areas that the Commission must address.

Phasing is the process of breaking down a large project into several smaller projects. I highly recommend developing an expandable system that is built in phases, starting with a pilot as discussed later in my testimony.

Q. Briefly discuss Step 2: Commence Physical Data Loading (Integration with Data Sources).

A. Data loading is the combined physical actions of New Hampshire utilities and non-utilities as they upload Energy Data to the SB284 Data Platform, based on agreed standards and permissions, resulting in the creation of the "common base of energy data" stipulated in RSA 378:51.

Data loading, also called data uploads and Data Source integration, is guided by the data requirements of the Use Cases supported. Different Use Cases may require additional data, although many Use Cases will reuse the same data if it can be sorted, filtered, aggregated or related differently in order to meet the data requirement of their Application. Data uploads are the essential step in executing an External Integration Strategy.

Q. Briefly discuss Step 3: Commence Standardized Integrated Data Sharing with agreed External IT Application(s).

A. This step requires the development of an API. The result of this step is Standardized Integrated Data Sharing to an agreed set of stakeholder External Application.

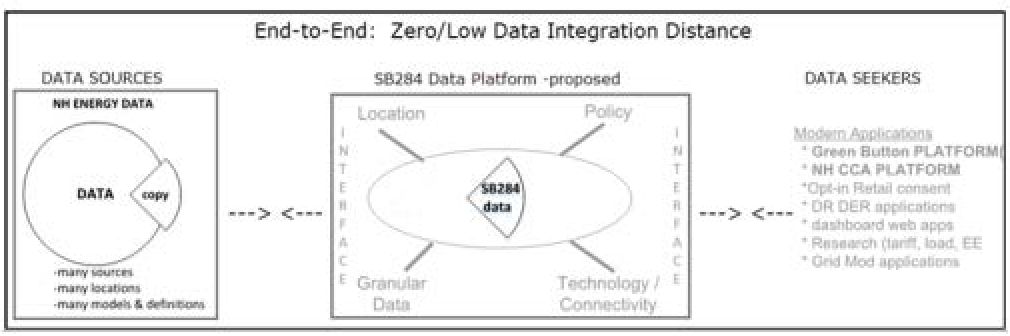
F. Data Platform Section Summary

Q. Please summarize the Data Platform and your conclusion.

A. The Data Platform should be the physical asset made of hardware and software that exists as New Hampshire's centralized data access point used and is trusted by energy stakeholders and third parties who create Applications and Platforms that require a common base of energy data. The Data Platform is the physical implementation of an External Integration strategy recommended as the most cost effective way to eliminate Data Integration Distance that causes friction and inefficiencies in the New Hampshire energy economy.

The proposed Data Platform should implement a cost effective External Integration strategy to integrate Data Sources and Data Seekers. Energy Data locked inside numerous legacy data systems will become, once uploaded into the Database using NEEDS as a Common Model, part of the Standardized Integrated Data Sharing and all its benefits. Publishing the APIs developed for the SB284 Platform will allow Data Seekers and Applications to discover and use the standardized data after passing the requisite permissions and qualifications.

Figure: 10: End-to-End: Zero/Low data Integration Distance



In conclusion, Figure 10 ("End-to-End: Zero Data Integration Distance") shows the SB284 Data Platform as a strategic asset that gives New Hampshire a long-term targeted process for Data Sources to migrate towards. This will be especially important as the New Hampshire regulated utilities move forward with grid modernization plans approved by the Commission in coming years. The SB284 Data Platform also provides New Hampshire a targeted process for other stakeholders, as both Data Seekers and potential Data Sources to migrate towards, where future Use Case Applications are targets. The External Integration Strategy (where the SB284 Data Platform is in the center) leads to robust data sharing and less friction in the system. When Standardized Integrated Data Sharing is achieved, then the SB284 Data Platform's accessible "common base of energy data" will enable stakeholders to create External Applications and IT

Systems (seen on the right side of Figure 10) that can digitally transform New Hampshire's energy economy.

IV. Risk and Governance

A. Risks

Q. How do you suggest avoiding risks for a SB284 Data Platform?

A. All projects hit road blocks that can be overcome with proper management and skill. However, an area that is key to prevent failure is to ensure that the business requirements are well defined, understood, and not in conflict. Well-designed and effective oversight processes are key to avoiding risks on the way to success.

Q. What are the key risks and your proposed mitigation strategies?

A. I identify six areas of potential risk that can and should be mitigated by being addressed early in the process.

1. Failure to design a good data model. A more robust data architecture may increase the initial costs of building the data platform. But, if designed well, based on comprehensive business analysis, can provide a lower risk cost effective path to implement additional use case functionality.

2. Risks of redesigning existing IT Systems. Risks, including delays and high costs, are often incurred by customers when their utility is forced to redesign an IT system that has already been designed, built, tested, deployed, and in use. Poor data quality is often a contributing factor in the decision to redesign an IT system, including an energy data sharing system.

3. Failure to plan strategically. Failing to design the data platform to meet the future realities of the grid will dramatically reduce its usefulness, create risks and added costs to modify or create new data sharing systems, and overall increase the risk of technology obsolesce. Designing the data platform with the capacity to maintain and share granular data provides future-proofing and risk mitigation.

4. Failure to make important design decisions early on. It is critical to develop an awareness of future potential requirements, early on, prior to designing the data platform. Design decisions early on, such as adoption of underlying data models, are beneficial in designing the system in such a way that future phases can be added cost effectively, as necessary and appropriate.

5. Failure to implement in phases: With proper planning External IT Systems such as the SB284 Data Platform can be implemented in phases. Phases are recommended as a means of managing costs and risk when designing and deploying a technology platform.

6. Failure to manage the SB284 Data Platform as an enterprise IT project. The platform should be managed like a traditional IT project including project planning, project management, and oversight of a documented System Development Life Cycle.

B. Governance

Q. To what extent is it important and/or necessary for the Commission to address issues related to the governance of the data platform?

A. It is essential. As I stated above, oversight is key to risk mitigation. In addition, SB284 explicitly requires the Commission to address the issue of "governance" in this adjudicative proceeding. The relevant language is codified as RSA 378:51, II(a). Even if the General Court had not issued such explicit instructions, the public interest requires the Commission to specify how the platform is to be governed.

Q. Why do you say that?

A. Because of the parameters the General Court established for the data platform. SB284 assigns to the state's electric and natural gas utilities the responsibility to "establish and jointly operate" the platform. The language is codified as RSA 378:51. But the provision is actually an instruction to the Commission, to "require" these utilities to execute this task, subject to the determinations the Commission must make in this adjudicative proceeding. Although I am not a lawyer, I assume the Commission will be mindful of the findings made by the General Court in the first section of SB284, which can also be found as section 1 of chapter 286 of the 2019 New Hampshire Laws. The General Court found that "safe, secure access to information about their energy usage" is important for consumers if New Hampshire is to accomplish the purposes of the electric industry restructuring process that was kicked off in 1996, to implement the state's energy policy as codified by RSA 378:37, and "to make the state's energy systems more distributed, responsive, dynamic, and consumer focused." Some of this energy policy and the goals of the SB284 Data Platform can be contrary to the interest of the utilities. With the utilities tasked with the establishment and operation of the platform, these legislatively-adopted policy objectives will not be realized unless the utilities are subject to robust and independent oversight at every step of the process.

Q. Why is robust and independent oversight essential?

A. As I have already testified, we are dealing with investor-owned utilities here. They are profit-maximizing firms whose management is obliged to seek as much return on investment for shareholders as possible. Both throughout their histories and now, the path to maximizing shareholder return involves making capital investments on a prudent basis and putting those investments into rate base. The data platform will inevitably constrain the utilities' ability to deploy capital in that fashion, because – as the General Court acknowledged in its findings – access to data will allow customers to assume more control over their energy usage and increase their reliance on unregulated, third-party service providers. That's what the General Court meant when it referred to making energy systems "more distributed, responsive, dynamic, and consumer focused." This process of reform and change can only come at the expense of utility opportunities to make their own investments, and provide their own services, as opposed to allowing customers to look to other providers or even just control their own energy usage. Essentially, the utilities have a conflict of interest here, and independent oversight is the cure.

Q. Are you suggesting that the utilities would not act in good faith if they were allowed to build and operate the data platform without oversight, beyond the regulatory constraints the Commission has historically provided?

A. No, not at all. I've been working with Unitil on data sharing since 2017, when the Company agreed to undertake such efforts with the OCA and Commission Staff as a part of the settlement of the rate case they filed the previous year, Docket DE 16-384. I was an active participant in the Grid Modernization Working Group that drafted a detailed report filed in IR 15-296, the grid modernization investigation. And I have been in regular and active communication with representatives of the utilities since this docket was opened last year. These contacts leave me convinced that each utility employs thoughtful and well-intentioned experts who are committed to helping their companies do more to facilitate customer and third-party access to usage data.

But the utilities' natural incentive to resist progress of this sort is still there. Independent oversight of the data platform, if done correctly, would have the effect of giving the advocates of data access within the utilities a powerful basis for arguing internally for the creation and operation of a truly consumer-empowering data platform. Basically, I am proposing to use governance to inoculate utility management from claims that, by creating a platform that is optimal from the perspective of consumers, management is being insufficiently attentive to the shareholders' desire for maximum return on investment. Armed with that sort of immunity, I am certain the relevant experts inside the utilities can do an excellent job of delivering a platform that will serve New Hampshire well and become a template for what other jurisdictions should build as well.

Q. Why isn't the usual Commission oversight of utilities adequate to this task, especially given that the Commission is tasked with the "general supervision" of every utility pursuant to RSA 374:3?

A. Traditional regulatory oversight is not adequate to the task. Governance in this context requires active oversight of the planning, development, and operation of the platform, in something close to real time. Typically, Commission oversight is an after-the-fact phenomenon, with utility expenditures and decision making examined for their prudence. It is important to keep in mind that what we are talking about here is very different from traditional utility activities. What consumers will get via the platform is very different from the safe and reliable delivery of electricity and natural gas that these utilities currently provide.

Q. How should the Commission approach the question of what governance mechanism is the right one?

A. The Commission should keep in mind that governance tasks fall into two distinct categories. The first has to do with the overall design of the platform. To the extent unresolved by the Commission's decision or decisions in this docket, the governance process must address questions like how the platform will be designed, what services and functionalities it will provide, what aspects of it will be developed by the utilities and what aspects will be provided by outside contractors, et cetera. The second category has to do with the operation of the platform. This involves questions like change management, software updates, and dispute resolution. The point here is not to offer a detailed definition of these two categories but, rather, to differentiate between planning and designing the platform as opposed to operating it. Each of these two realms requires a distinct approach to governance if the platform is to be effectively and efficiently deployed.

Q. What general approach do you recommend for governance in the areas of planning and design?

A. Ideally, because the utilities are tasked with establishing and jointly operating the platform under the statute, they would play no direct role in governance at this meta-level. Instead, they would essentially act as a service provider, accountable to a governing body that would be comprised of ratepayer representatives as well as representatives of third-party service providers, municipalities (in light of the development of Community Power Aggregation pursuant to RSA 53-E, something the SB284 findings explicitly references), and perhaps the Commission's electric division.

Alternatively, the Commission could adopt an approach similar to the one endorsed in the PUC's May order in the grid modernization investigation, Docket IR 15-196. It is also the approach that drives the development of plans for implementation of New Hampshire's Energy Efficiency Resource Standard (EERS). In both of those realms, the utilities are themselves considered stakeholders and thus participate in oversight bodies as equals alongside ratepayers, service providers, and others with legitimate interests – although, of course, the utilities do not (and should not) enjoy voting majorities on any such deliberative bodies.

Q. Do you have a specific recommendation?

A. Yes. I recommend the creation of a nine-member stakeholder governance board, comprised of the Consumer Advocate or his designee (to represent the interests of residential customers), a representative of small commercial customers, a representative of large commercial customers, two members of the Commission Staff, two municipal representatives, and two representatives of firms that provide energy-related services to consumers that depend on access to data. I would task the Commission with appointing all of these representatives other than the Consumer Advocate (or designee). Alternatively, the size of the stakeholder governance board could be increased to 12 voting members with a representative of Eversource, Liberty, and Unitil each given one vote.

Q. Should the Commission determine how this body will operate?

A. Not at this time. In my opinion, it would be better to allow this body to convene and work to decide for itself how it will operate, presumably according to bylaws or some similar governance document the body would adopt. I would, however, make these organizational determinations subject to the Commission's approval. Given the nascent state of the data platform at this juncture, it does not make sense to dictate how the body will operate until it has had a chance to convene and consider these questions itself.

Q. Would there be any role for the utilities if they do not have voting membership in the body?

A. Yes. The Commission should make clear it expects the utilities to attend meetings of the body, to provide such information and advice to the body as it might require. However, I would specify that the body can seek outside assistance, including consultants as needed, subject to Commission approval.

Q. Should the Commission rely on this process to provide day-to-day governance of the platform?

A. No. For this purpose I would create a platform operations committee that would be comprised of three utility representatives (one each from Eversource, Liberty, and Unitil), three representatives of third-party service providers reliant on the platform for data, and a tie-breaking representative of the Commission Staff. I would allow each utility to appoint its own representatives and have the rest of the operating committee appointed by the Commission.

Q. How should the platform operations committee do its work?

A. The key here is nimble and efficient decision making. The committee should be responsible for operationalizing the initial and ongoing requirements established under the governance body. A key responsibility would be the review of changes to the technology, implementation, and functional requirements of the platform quickly, as the need for such changes arises in real time. There is also likely to be a need to resolve disputes in the event that platform users encounter obstacles or difficulties. It would make sense to allow the platform committee to authorize subcommittees to make decisions quickly, subject to review by the entire committee. Disputes within the committee should be brought to the governance board for resolution. If there is a need to resolve conflicts between the Committee and the Board these would go to the Commission. As with the stakeholder governance board, I would allow the operations committee to design its own operating rules in the first instance, subject to approval by the governance board and the Commission. That makes more sense than attempting to pin down all the details now.

Q. A different approach to governance could involve the creation of a new and independent business entity to develop, own, and operate the platform. Why didn't you recommend this approach?

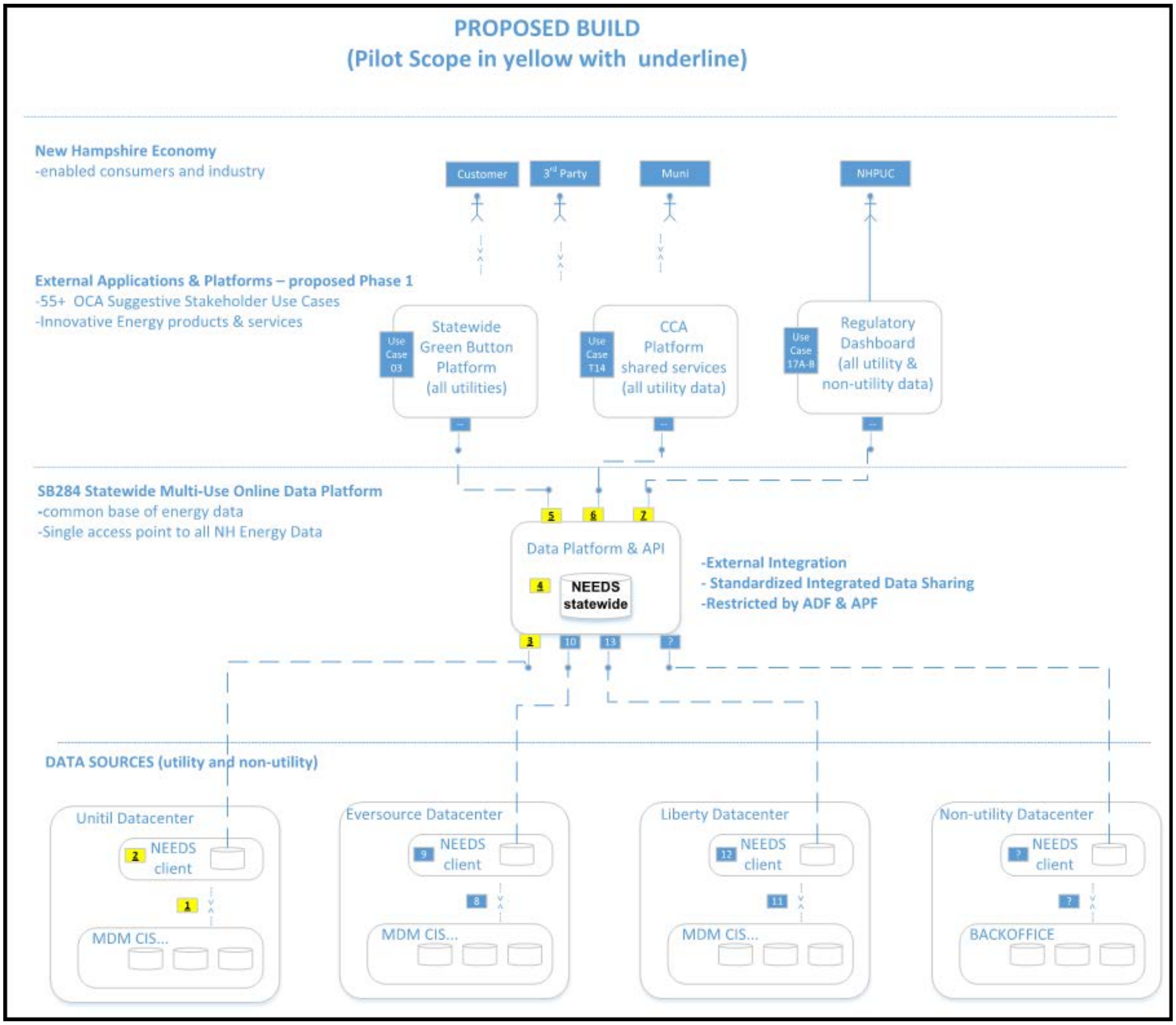
A. I agree that a new and independent entity would be an effective way of assuring that the utilities do not dominate the platform or, perhaps, design and/or operate the platform in a way that is only in the best interests of utility shareholders as opposed to customers and other stakeholders. I note that SB284 does not rule out such an approach even though it tasks the utilities with establishing and jointly operating the platform. The utilities could certainly establish and operate the platform under contract to a new and independent entity. However, this approach is likely to introduce some challenging complexities, especially when it comes to cost recovery, given that SB284 states only that the *utilities* may recover their prudently incurred costs from customers. While I would not object if the Commission were ultimately to adopt the 'independent entity' approach, I believe the same general objective of platform independence can be achieved via reliance on the two governance bodies the establishment of which I am proposing.

V. Proposed Build – Conceptual Five-Year Blueprint

Q. Looking back to Figure 1 "Platform Integration" that began your testimony, what is your proposed build relative to Box 3A Data Platform, Box 3B Data Sources and Box 3C Applications / Data Seekers?

A. My proposed build is shown in Figure 11, "Proposed Build." It is a high level conceptual blueprint. My proposed build would integrate Data Sources (bottom level of diagram), with Data Seekers which are External Application and the New Hampshire energy economy (top two layers of the diagram), by implementing an External Integration using SB284 Statewide Multi-Use Online Data Platform (center level of diagram). Figure 11 suggests there are 13 numbered tasks that would need to be completed by various parties. When built, the platform would support the three proposed Phase 1 Use Cases discussed in my testimony: (1) Statewide Green Button Platform, (2) CCA Application/Platform, and (3) Centralized Regulatory Dashboard. The proposed build would also support statewide sharing of non-utility data as illustrated in the bottom right area of Figure 11, "Non-Utility Datacenter. Finally, the diagram illustrates a potential scope of work for a Design Pilot to formally design parts of key components. In the diagram these Design Pilot task are identified as tasks 1, 2, 3, 4, 5, 6, and 7 (highlighted and underlined in yellow).

Figure 11: Proposed Build (Including Design Pilot Scope)



Q. Is your recommendation to build the entire platform immediately upon the Commission's order following the hearing in this docket?

A. No. Figure 11 is a conceptual blueprint for achieving the vision and benefits of an External Integration strategy that I discussed in the Vision section. The diagram is informational and could be followed in a stepwise, prudent fashion over a period of two to five years, to bring the benefits of Standardized Integrated Data Sharing to New Hampshire. While high level and with many details yet to be determined, the diagram reflects ideas that have been developed, analyzed and to varying degrees vetted with in-state and national experts as discussed earlier. Should the Commission decide there is merit in considering some version of the blueprint suggested in Figure 11, I would recommend the initiation of a formal IT Project Plan, following a SDLC. The Project Plans next steps would be formal design and testing of some of the key components identified in the diagram as 1, 2, 3, 4, 5, 6, and 7 (highlighted and underlined in yellow). Once the design of the key components is completed, planning for building the system can start to be developed.

VI. Recommendations and Analysis

A. Recommendations

Q. In light of the testimony you have already given, please list and briefly describe your recommendations to the Commission.

A. My recommendations are:

1. Create working group(s) that will develop two key sets of policies: Data Access Framework (DAF) and Data Privacy Framework (DPF). Both of these umbrella policies control data access and privacy to protect customers. The DAF and DPF are shown in Figure 11 "Proposed Build" in the Proposed Build and labeled "[SB284 is] Restricted by DAF and DPF."

2. Develop final Use Cases, including Sequence Diagrams and Use Case data needs for the proposed three Use Cases that the SB284 Data Platform will support and enable in Phase 1: (1) Green Button Platform, (2) Community Choice Aggregation, and (3) Regulatory Dashboard.

3. Develop Logical Data Model (NEEDS). This recommendation is closely related to the Design Pilot discussed below. It is listed as a separate recommendation to account for any differences that may arise in governance of required areas of expertise between the two recommendations. An example of a potential difference between developing NEEDS and conducing a Design Pilot is that subject matter experts from non-utility areas are required in the design process of developing NEEDS.

4. Initiate a Design Pilot to design and test two key aspects of New Hampshire's model prior to moving forward with a full buildout. The design pilot requires Unitil create a physical environment that will be used to develop and test specific portions of numbered tasks 1 to 7 in Figure 11. The Design Pilot would design and test the feasibility of using a Common Data Model, NEEDS, to facilitate statewide Standardized Integrated Data Sharing that includes all New Hampshire investor owned energy utilities. The design pilot will be a milestone moving toward building the actual platform.

a. Create the Database defined by a Physical Data Definition adhering to NEEDS-client (numbered task 2, and data mapping and data loading from legacy systems (numbered task 1).

b. Physical Database defined by a Physical Data Definition adhering to NEEDS-statewide (numbered task 4) and uploading of Unitil Energy Data to the Physical Database (numbered task 3).

c. Preliminary work and testing (including datasets created by the Data Platform) of an API endpoint for a statewide Green Button Platform producing certified file (numbered task 5).

d. Preliminary work and testing (including datasets created by the Data Platform) of an API endpoint for a CCA Application (numbered task 6).

e. Preliminary work and testing (including datasets created by the Data Platform) of an API endpoint for a Regulatory Dashboard (numbered task 7).

5. Create the two governance bodies I describe in the Governance section of my testimony, ideally soon enough so that they can have meaningful input into the creation of the Design Pilot.

Q. Why do you recommend Unitil for the Pilot?

A. Unitil is recommended based on three factors: (1) existence of Advanced Metering Infrastructure which would allow testing of granular data and TOU measurements across all customer classes, (2) Unitil's affiliates, UES and Northern, are Data Sources for electric and gas usage data which would allow analysis and testing of multi-fuel Energy Data, (3) Unitil's knowledge of the DE 16-384 efforts, including early work on the Logical Data Model.

Q. Is the purpose of the design pilot to create a working data sharing platform?

A. No. the purpose of the pilot is to evaluate and prove out critical concepts and would not require complete production versions of these tasks. As discussed above, the two critical concepts are (1) use of a Common Data Model, and (2) use of a centralized platform (to perform data sharing).

B. Design Pilot Recommendation & RSA 378:52, I

Q. Why is planning a key issue that the Commission should address?

A. Creating and launching a Data Platform involves coordinating thousands of complex, business, technical and operational aspects into an IT plan. In the case of an IT Platform, the plan needs to be structured based on principles called the System Development Life Cycle.[56] Regardless of what IT plan New Hampshire uses, a plan needs to exist, be overseen, and executed in a way that ensures none of the stages of the SDLC are omitted. For example, failure to design a platform prior to building a platform would be imprudent, risky, and illogical. Yet, SDLC steps do get skipped[57] and can contribute to project failures. In my opinion, a centralized SB284 IT project plan that follows the SDLC, where a senior lead who is authorized, accountable, and represents the interests of the stakeholder governance board, is needed in this docket at this time.

Q. What does RSA 378:52, I stipulate?

A. This provision of SB284 stipulates that the utilities shall "design and operate the energy data platform to provide opportunities for utilities, their customers, and third parties to access the online energy data platform and to participate in data sharing." This clearly refers to the process of system design.

Q. How should the Commission interpret this requirement for the utilities to embark on a process of system design?

A. A common definition of design is

> "a plan or specification for the construction of an object or system or for the implementation of an activity or process, or the result of that plan or

[56] The Software Development Life Cycle SDLC is defined and discussed in OCA's Scoping Comments of March 11, 2020 (Tab 24) at 10.

[57] Failure to perform SDLC steps adequately, such as requirements analysis, design, development and testing, implementation, documentation, and evaluation were noted during my outreach discussions.

> specification in the form of a prototype, product or process. The verb *to design* expresses the process of developing a design"[58]

Notice that the definition expresses "design" to be the "process of developing" (for example developing a prototype) directly applies to the SDLC. It is directly relevant to SDLC based planning. The "D" in SDLC stands for Development. The Design Pilot will develop parts of a prototype of New Hampshire's future data sharing model. .

Q. Please describe the "Design Pilot" you have recommended.

A. The "Design Pilot" is a group of design tasks that will produce a prototype of critical aspects of the proposed platform. The Design Pilot will transform the immature concepts that emerged from the DE 16-384 DWG (i.e., ideas on paper) to a mature design that can be tested, and analyzed by other stakeholders. The Design Pilot would operate in in a highly iterative back and forth fashion (i.e., design/test/re-design/re-test, et cetera) and produce working prototypes of key elements of the platform. The prototype would not produce a live system for New Hampshire to access. It could however provide designs that, when provided to vendors, result in them performing less work and thereby reduce costs significantly.[59] This could be a very significant asset when conducting a competitive RFP processes, and in determining what work needs to be done, and what work has already been developed by vendors awarded contracts building the live statewide platform.

[58] https://en.wikipedia.org/wiki/Design.

[59] Anecdotally, I am informed that Smart Meter Texas cost $10 million to build, including the design.

Q. Who would perform the work of the Design Pilot?

A. My recommendation is that a significant portion of design and architecture work, including the work started in DE 16-384 DWG, be performed by a New Hampshire Pilot team including in-house experts from Unitil. This includes further development of the Logical Data Model. An outside consultant can be strategically used and in specialized areas. Performing a significant portion of the work using the in-house resources of parties to this docket would promote and develop in-house knowledge and ownership of this exciting transformative project to create a Data Platform. Knowledge gained during this phase will be useful when the platform is expanded to other utilities.

Q. Why are you recommending a pilot and not a full build out of the statewide platform immediately?

A. I believe the risks of immediately moving forward with a full platform build exceed the benefits. I also believe the benefits presented in conducting a pilot are substantial, low-risk, and affordable. I will make three points supporting this recommendation.

First, a small work pilot could break the ice entrenched positions and lead to compromise in real time. In my experience in IT I have found experts working on technical teams tasked with large goals, assuming that team properly orchestrated founded in trust and good faith, often leave their employer hats at the door and strive to succeed, hitting the common goal.

Second, New Hampshire's opportunity to build on 18-24 months of foundation work, vetted and nationally recognized, to develop a prototype centralized data model and platform, is an opportunity with tremendous upside, and limited downside potential. I am not sure the same can be said of undertaking of the tremendous task of designing, developing, building and

100

operations of a statewide platform at this time. If the Design Pilot is successful, the cost to build the platform will significantly reduce the cost to build the final system that meets New Hampshire's data related grid modernization goals for years to come. A prototype, can reduce expensive work of a vendor designing a system and reduce the RFP project scope to a somewhat more proven set of IT tasks, and perhaps a different category of vendors that is tasked with operationalizing the prototype design. If the pilot fails to produce a prototype, the cost will be small compared to the cost of a full high risk statewide build. However, even under failure, success and benefits can emerge, knowledge gained, team work culture established and the "right" path found – hopefully a path that does not repeat the mistakes of other states seen over the past 5-10 years.

Third, the data sharing space is in a state of evolution. The current data sharing models that now exist in states today may not be the model New Hampshire wants to definitively replicate in 2021. There are significant issues with the utility-by-utility approach, some of which I have discussed in the section on influence from other states.

In conclusion, before New Hampshire picks its path and model, it makes strategic sense to wait, assess what the merits are of new models, and consider the advantages of a centralized platform. I am aware of two options available to provide such insights. The two options are (1) the New York DER pilot, and (2) the Design Pilot that will test a Common Data Model operating on a statewide basis, as proposed in my testimony. It is my opinion that New Hampshire should hold off on a major RFP until the merits of a centralized platforms can be analyzed. I believe the cost of the proposed pilot are significantly outweighed by the potential benefits.

101

VII. Summary and Conclusion

Q. What is your final Summary and Recommendation?

A. RSA 378:50 defines "data sharing" as "providing data and accessing data provided by others." RSA 378:51 calls for establishing a platform that "consists a common base of energy data for use in a wide range of applications and business uses." My testimony offers a vision for Data Sharing that complies with the requirements of sections 50 through 54 of RSA 378. That vision is the creation of a dedicated independent Data Platform that becomes the centralized data access point used and trusted by energy stakeholders and third parties who create Applications (energy services and products) that digitally transform New Hampshire's economy. The vision is long-term, achievable in a prudent step-wise targeted fashion, and will lead to reductions in costly systemic invisible friction that blocks innovation and realization of New Hampshire's energy goals.

The invisible friction is a symptom of underlying entrenched problems. Reducing the friction, which is the ultimate goal of my vision and my recommendations, requires the underlying problems to be identified and solved. I have identified the two underling problems that are the cause of systemic friction in New Hampshire's energy economy. They are (1) poor data quality, and (2) poor data access. I have also provided a new metric called "Data Integration Distance" that makes the invisible friction visible.

Solving the underlying problem, Data Integration Distance, requires a strategy and a means of implementing the strategy. My testimony proposes both the strategy and the means of implementing the strategy. I have proposed a strategy called External Integration that is widely used in many industries include the electric distribution industry. And I have proposed an

102

implantation of the strategy using a centralized Data Platform shown in Figure 11. The strategy (External Integration) and the implementation of the strategy (centralized Data Platform) have been analyzed and vetted locally and national with experts, and can be tested in a pilot and expanded, once the strategy is confirmed, to a state wide level over the next two to five years in a targeted fashion.

The Data Platform is a green field opportunity for the Commission to consider. It is a physical asset (hardware, software) that does not replace an existing physical asset already in rate base, or already being paid for. No such platform or legacy data sharing platform exists that can provide Standardized Integrated Data Sharing. It is in the public good to provide Energy Data to the New Hampshire economy to kick start innovation envisioned in grid modernization. The General Court has directed the Commission, via RSA 378:51, I(a), to see that New Hampshire's utilities "provide a common base of energy data for use in a wide range of applications and business uses." This is good time to think strategically about where New Hampshire want to be five years from now.

Appendix 1: STE Presentation and Handout
Appendix 2: OCA Suggested Use Cases with Brief Description

103