MARKET
Market and Retail-rate Knowhow for the Energy Transition

Moving Beyond Direct Load Control: A Maturity Model for Realizing the Promise of Demand Flexibility

January 2025

Josh Schellenberg, Berkeley Lab affiliate
Natalie Mims Frick, Berkeley Lab

NREL Transforming ENERGY
Argonne NATIONAL LABORATORY
BERKELEY LAB

Disclaimer

Copyright Notice

Moving Beyond Direct Load Control: A Maturity Model for Realizing the Promise of Demand Flexibility

Prepared for the
U.S. Department of Energy
Office of Energy Efficiency and Renewable Energy

Principal Authors
Josh Schellenberg
Natalie Mims Frick

Ernest Orlando Lawrence Berkeley National Laboratory
1 Cyclotron Road, MS 90R4000
Berkeley CA 94720-8136

January 2025

The work described in this study was funded by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy under Lawrence Berkeley National Laboratory Contract No. DE-AC02-05CH11231.

Acknowledgements

The work described in this study was conducted at Lawrence Berkeley National Laboratory and supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy under Contract No. DE-AC02-05CH11231.

The authors thank Paul Spitsen, Michelle Boyd, Stacey Miller, JP Carvallo and Robert Margolis, U.S. Department of Energy, for supporting our work, and the following experts for reviewing this report (affiliations do not imply that their organizations support or endorse this work):

Philip Chao and Paul Heitmann, New Jersey Board of Public Utilities
Robin Maslowski, Trillium Energy Consulting
Nicholas Papanastassiou, EnergyHub
George Twigg, New England Conference of Public Utilities Commissioners

Table of Contents

Table of Figures

Executive Summary

Despite the widespread adoption of advanced meters and increased need for demand flexibility[1] (DF) due to load growth and increasing renewable energy supply on the grid, utilities have generally not moved beyond traditional direct load control (DLC) programs to realize the promise of DF. Based on an in-depth review, Murphy et al. (2024) finds that most DF efforts focus on reducing demand during summer peaks driven by space conditioning consumption, similar to the findings of a 2006 Federal Energy Regulatory Commission (FERC) survey. Since 2007, the number of advanced meters has increased by over 112 million, from 4.7% penetration in 2007 to 72.3% in 2022 (FERC 2024). Prior to this growth, FERC estimated that the total potential demand response[2] (DR) resource contribution from existing programs and rates in 2006 was 37.5 GW (FERC 2006), but recent data shows that nationwide DR capacity may currently be lower. While many factors have introduced uncertainty into assessing multi-year DF trends throughout the country, it is clear that DF programs have not grown or changed substantially, even though DF potential has increased as utilities have rapidly deployed advanced meters and consumers have adopted a wide range of connected devices.

With surging electricity load growth from data centers, manufacturing facilities, buildings and transportation, DF is urgently needed to mitigate capacity constraints and power interruptions. Given the slow progress with DF programs recently, utilities and regulators are seeking novel approaches for driving DF progress. Therefore, Berkeley Lab has developed the DF maturity model (Figure ES-1), based on well-documented industry best practices and barriers to broader DF adoption.

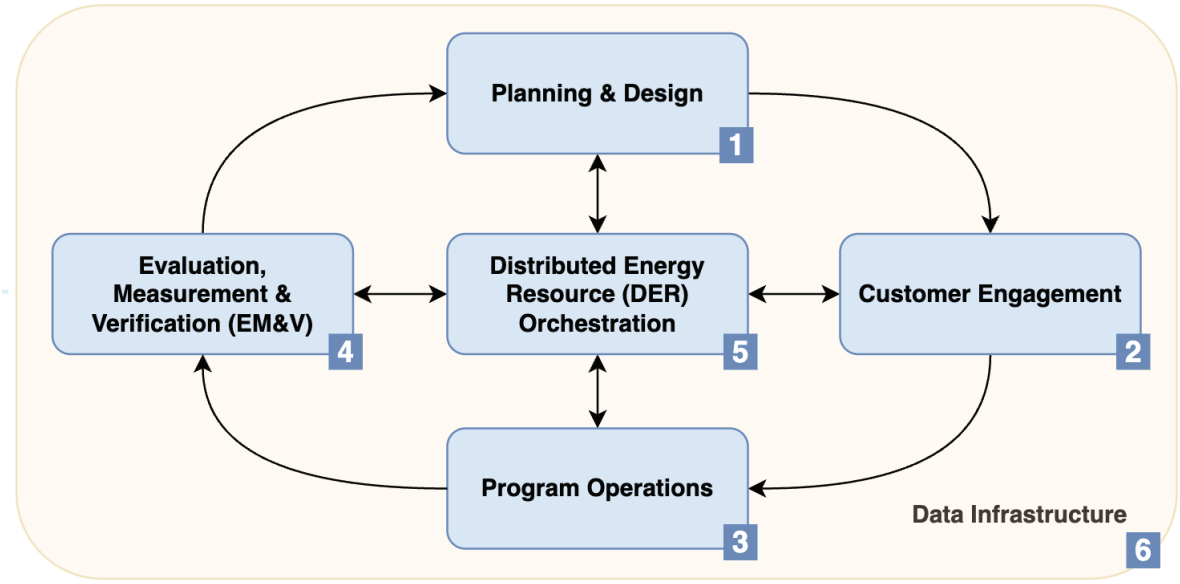


Figure ES - 1. Berkeley Lab Demand Flexibility Maturity Model

[1] Demand flexibility includes any type of flexible behind-the-meter Distributed Energy Resource (DER), including solar, storage, electric vehicles, and traditional demand response (DR). It is defined as "the capability provided by on-site DERs to reduce, shed, shift, modulate, or generate electricity" (DOE 2021).

[2] DR and DF are used interchangeably in this report for the purposes of assessing progress.

This detailed, practical guidance based on a wide range of lessons learned fills a gap in the literature by providing an extensive tool to help utilities and regulators realize the promise DF. Importantly, the maturity model primarily applies to retail DF programs run by utilities, with involvement of third-party technology providers and DER aggregators. The model could also apply to dispatchable time-varying pricing programs (e.g., Critical Peak Pricing), though it does not include key pricing-related capabilities such as rate design. States and utilities can also adapt the maturity model to their jurisdiction as needed.

The Berkeley Lab DF maturity model scores utility programs on 34 DF capabilities, organized into the six categories in Figure ES-1. Similar to utility capability maturity models for wildfire mitigation (Office of Energy Infrastructure Safety 2020), cybersecurity (DOE 2022) and smart grid (Carnegie Mellon 2018), the DF maturity model is a framework for assessing current capabilities and developing a plan for improvement. Similar to the California Wildfire Mitigation Maturity Model (Office of Energy Infrastructure Safety 2020), each DF capability is scored at one of five possible levels of maturity:

1. Below expectations
2. Meets minimum expectations
3. Beyond minimum expectations, but not consistent with best practice
4. Consistent with best practice
5. Improvement over best practice

Industry literature, conference presentations, and reviewer feedback have informed maturity levels, including best practices and reasonable expectations of utility capabilities.

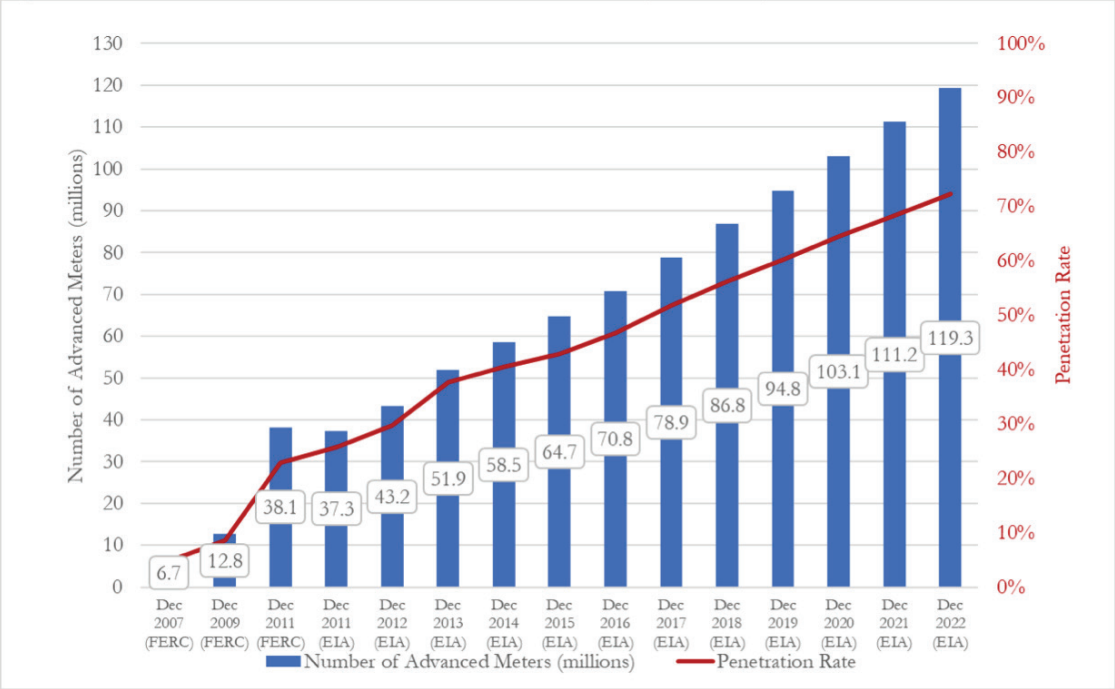
By measuring maturity levels across 34 DF capabilities, states and utilities can prioritize investment, identify gaps, and drive continuous improvement to realize the promise of DF and move beyond DLC. In general, the levels of maturity drive DF programs in the following direction:

- From inflexible and uninformed by grid needs and data (low maturity) to highly flexible and driven by grid needs and accurate, granular data (high maturity)
- From a limited understanding of performance (low maturity) to frequent, ongoing performance assessment that proactively identifies scaling constraints and continuous improvements (high maturity)
- From infrequent, reactive testing and iteration (low maturity) to proactive testing and monitoring of key leading indicators based on randomized tests, leading to rapid iteration including “failing fast” when necessary (high maturity)
- From a focus on enrollment as the primary success metric (low maturity) to expected customer lifetime value based on the projected net benefits by customer group (high maturity)
- From new pilots and initiatives that do not scale efficiently (low maturity) to a focus on driving value at scale, with close integration of DF program planning, operations, and external parties, including system operators and third parties as needed to scale efficiently (high maturity)

1. Introduction

Despite the widespread adoption of advanced meters and increased need for demand flexibility[3] (DF) due to load growth and increasing renewable energy supply on the grid, utilities have generally not moved beyond traditional direct load control (DLC) programs to realize the promise of DF. Based on an in-depth review of 148 programs throughout the United States, Murphy et al. (2024) finds that most DF efforts focus on reducing demand during summer peaks driven by space conditioning consumption. This finding is remarkably similar to a 2006 Federal Energy Regulatory Commission (FERC) survey, which concluded that the most common demand response[4] (DR) programs nearly 20 years ago were traditional peak shedding programs, most notably air-conditioning DLC aimed at peak load shedding.

The 2006 FERC survey also found that interest in DR programs was growing, and results from recent programs and pilots were encouraging, based on responses from 3,365 organizations in all 50 states representing every aspect of the electric delivery industry. However, this increased interest has not resulted in a significant change in the types of DF programs being implemented, despite growth of advanced meters that enable a wider range of grid services.



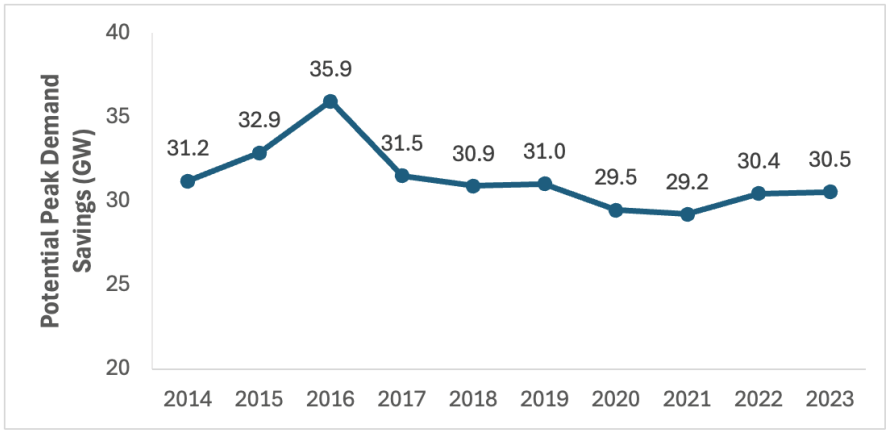
Source: FERC (2024)

Figure 1-1. Advanced Meter Growth in the United States (2007–2022)

[3] Demand flexibility includes any type of flexible behind-the-meter Distributed Energy Resource (DER), including solar, storage, electric vehicles, and traditional demand response (DR). It is defined as "the capability provided by on-site DERs to reduce, shed, shift, modulate, or generate electricity" (DOE 2021).

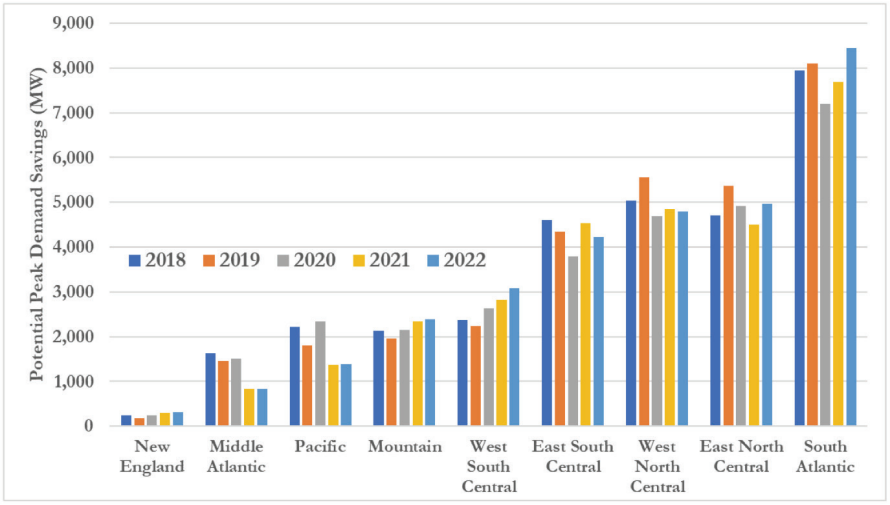
[4] DR and DF are used interchangeably in this report for the purposes of assessing progress, though DF is generally more associated with the broader range of grid services that DERs can provide.

Since 2007, the number of advanced meters has increased by over 112 million, from 4.7% penetration in 2007 to 72.3% in 2022 (Figure 1-1). Prior to this growth, FERC estimated that the total potential resource contribution from all existing DR programs and rates nationwide in 2006 was 37.5 GW (FERC 2006). Recent data from the U.S. Energy Information Administration[5] (EIA) shows that retail DR program capacity is lower, at about 30.5 GW, as measured by the potential peak demand savings (Figure 1-2) (EIA 2023). Furthermore, the EIA data shows that potential peak demand savings have not increased from 31.2 GW in 2014, after increasing to nearly 36 GW in 2016 and then declining. Regionally, FERC (2024) reports an annual increase in retail DR program potential peak demand savings in seven of the nine census divisions in 2022 (Figure 1-3), but consistent multi-year growth is not evident in any region.



Source: EIA (2023)

Figure 1-2. Retail DR Program Potential Peak Demand Savings in the United States (2014–2023)



Source: FERC (2024), based on Form EIA-861 data

[5] EIA started collecting retail DR program data from utilities via Form EIA-861 in 2013. The potential peak demand savings estimates do not include dynamic pricing programs, also known as time-based rates.

Figure 1-3. Retail DR Program Potential Peak Demand Savings by Census Division (2018 – 2022)

While changes in metrics, measurement methods, market participation rules, and other factors have introduced uncertainty into assessing multi-year DF trends throughout the country, it is clear that DF programs have not grown or changed substantially, even though DF potential has increased as utilities have rapidly deployed advanced meters and consumers have adopted a wide range of connected devices. As emphasized in Murphy et al. (2024), DF programs will need to evolve and grow to achieve a vision of DF that involves the provision of a wider variety of grid services, including using on-site DERs to reduce, shed, shift, modulate, or generate electricity.

Part of the slower-than-expected DF evolution may have been due to low load growth in the past 20 years. With surging electricity load growth from data centers, manufacturing facilities, buildings and transportation, DF is now urgently needed to mitigate capacity constraints and power interruptions. Given the slow progress with DF programs recently, utilities and regulators are seeking novel approaches for driving DF progress, which may include applying the Berkely Lab DF maturity model in their jurisdiction. Vast DF potential remains for existing behind-the-meter technology to alleviate grid constraints at a low cost. Industry best practices and impediments to DF progress are well-documented, including barriers to broader adoption of:

- DR (Delurey 2013);
- Integrated Demand Side Management (Potter et al. 2018);
- Grid-Interactive Efficient Buildings (DOE 2021);
- Virtual Power Plants (DOE 2023); and
- Third-party DF participation (see text box).

Berkeley Lab has integrated this knowledge base into the DF maturity model, summarized in Section 2 and provided in its entirety in the Appendix. Similar to the California Wildfire Mitigation Maturity Model (Office of Energy Infrastructure Safety 2020), each DF capability is scored at one of five possible levels of maturity. By measuring maturity levels across 34 DF capabilities, states and utilities can prioritize investment, identify gaps, and drive continuous improvement to realize the promise of DF and move beyond DLC. In combination with other metrics, states can also use the maturity levels as indicators of DF progress as part of utility financial incentives.

The remainder of this report includes a DF maturity model overview, references to further resources, and an Appendix with the full maturity model that can be adapted as needed.

Third-party DF Participation in the DR Auction Mechanism in California

Given the increased interest in third-party DF, the maturity model draws from well-documented recent learnings and challenges with third-party participation, most notably as part of the DR Auction Mechanism (DRAM) in California. DRAM started out in 2015 with the promise that, "For the first time, aggregations of customer-sited resources of at least 100 kW will have access to the state's electricity markets" (Walton 2015). It was an opportunity for third-party DR providers to participate in the wholesale market. However, after several years and over $100M allocated to the statewide DRAM procurement budget, the CPUC decided to close the program (CPUC 2024), citing an evaluation report which could not affirmatively find that DRAM met all six success criteria, particularly in the areas of performance and reliability (Schellenberg et al. 2022). The key issues identified by the evaluation were:

- Widely varying performance, and consistent underperformance in some cases, have significantly lowered the overall effectiveness of DRAM
- Lack of availability during critical hours, such as the August 2020 heatwave
- Given the above two issues, the additional system capacity (if any) that each third-party DR provider has delivered is highly uncertain and varies substantially
- Data errors and reporting inconsistencies, leading to over-compensation and a lack of confidence and transparency regarding overall DRAM performance
- Significant administrative burden for all parties involved

The DF maturity model builds from the learnings of DRAM by emphasizing that states and utilities closely monitor performance and availability during critical hours, ensuring that avoided costs are tied to demonstrably offsetting grid investments. The maturity model also includes standalone categories for DER orchestration and data infrastructure to highlight the importance of data accuracy and reducing administrative burden for all parties involved at high levels of maturity.

2. Demand Flexibility Maturity Model Overview

The Berkeley Lab DF maturity model scores utility programs on 34 DF capabilities, organized into six categories in Figure 2-1. Similar to the California Wildfire Mitigation Maturity Model, each DF capability is scored at one of five possible levels of maturity:

1. Below expectations
2. Meets minimum expectations
3. Beyond minimum expectations, but not consistent with best practice
4. Consistent with best practice
5. Improvement over best practice

For each capability, states and utilities can measure maturity levels to prioritize investment, identify gaps, and drive continuous improvement. In combination with other metrics, states can also use the maturity levels as indicators of DF progress as part of broader utility financial incentives.

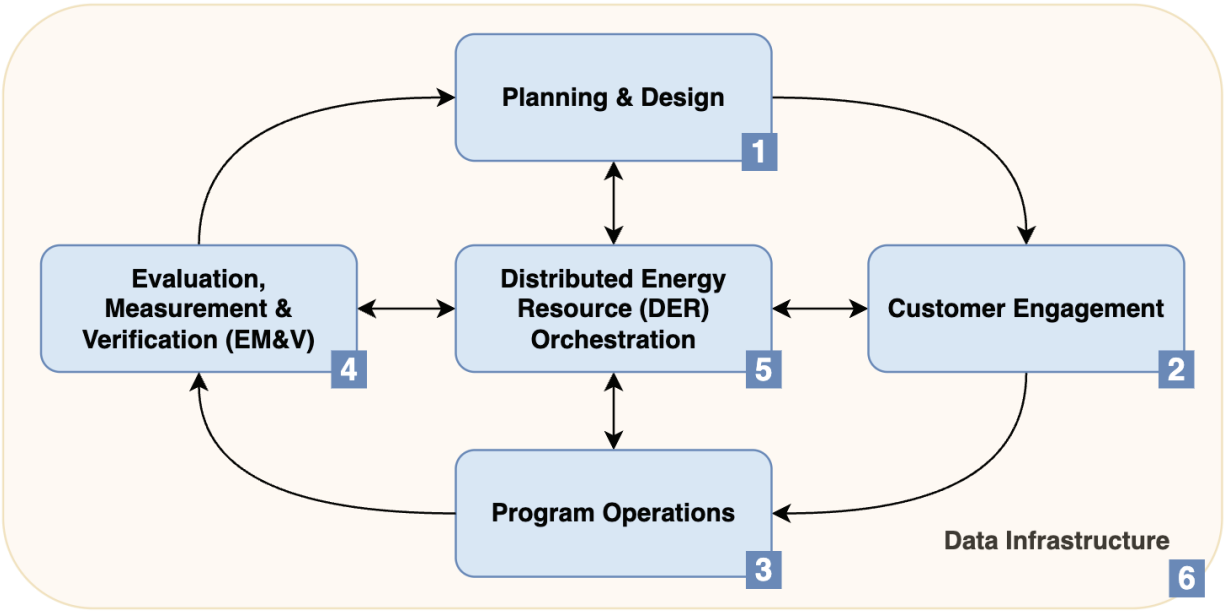


Figure 2-1. Berkeley Lab Demand Flexibility Maturity Model

Importantly, the DF maturity model is not necessarily a progression with a higher level clearly being more appropriate than a lower level for every jurisdiction. States and utilities can skip levels or stay at current maturity if that is what is cost-effective based on grid needs, market conditions, and the relative cost of meeting or exceeding best practices. It also may be cost-effective to have a mix of maturity levels within a DF portfolio. For example, a legacy air-conditioning DLC program may have a lower level of maturity on average, but the program can provide significant load shedding that is valuable under system peaking conditions. The same utility may also seek a high level of maturity for programs that deliver a wider range of grid services, including using on-site DERs to reduce, shift, modulate, or generate electricity.

2.1 DF Maturity Model Categories and Capabilities

Each of the six maturity model categories has 5 to 7 capabilities (Figure 2-2). Maturity levels are generalized, with limited use of terms specific to certain jurisdictions or technologies, so that the maturity model can apply to varying regulatory environments, market structures and types of grid services available. Nonetheless, states and utilities can adapt the maturity model as needed. In a few cases, there is some overlap between capabilities to ensure that maturity is tracked from more than one perspective when capabilities are interconnected (such as the role of third parties from the perspectives of Planning & Design, Customer Engagement and DER Orchestration). Finally, while the maturity model details the requirements for meeting a given level of maturity level for each capability, comprehensive implementation details are beyond the scope of this report. The remainder of this section summarizes each category and how each jurisdiction can adapt the model to their region.

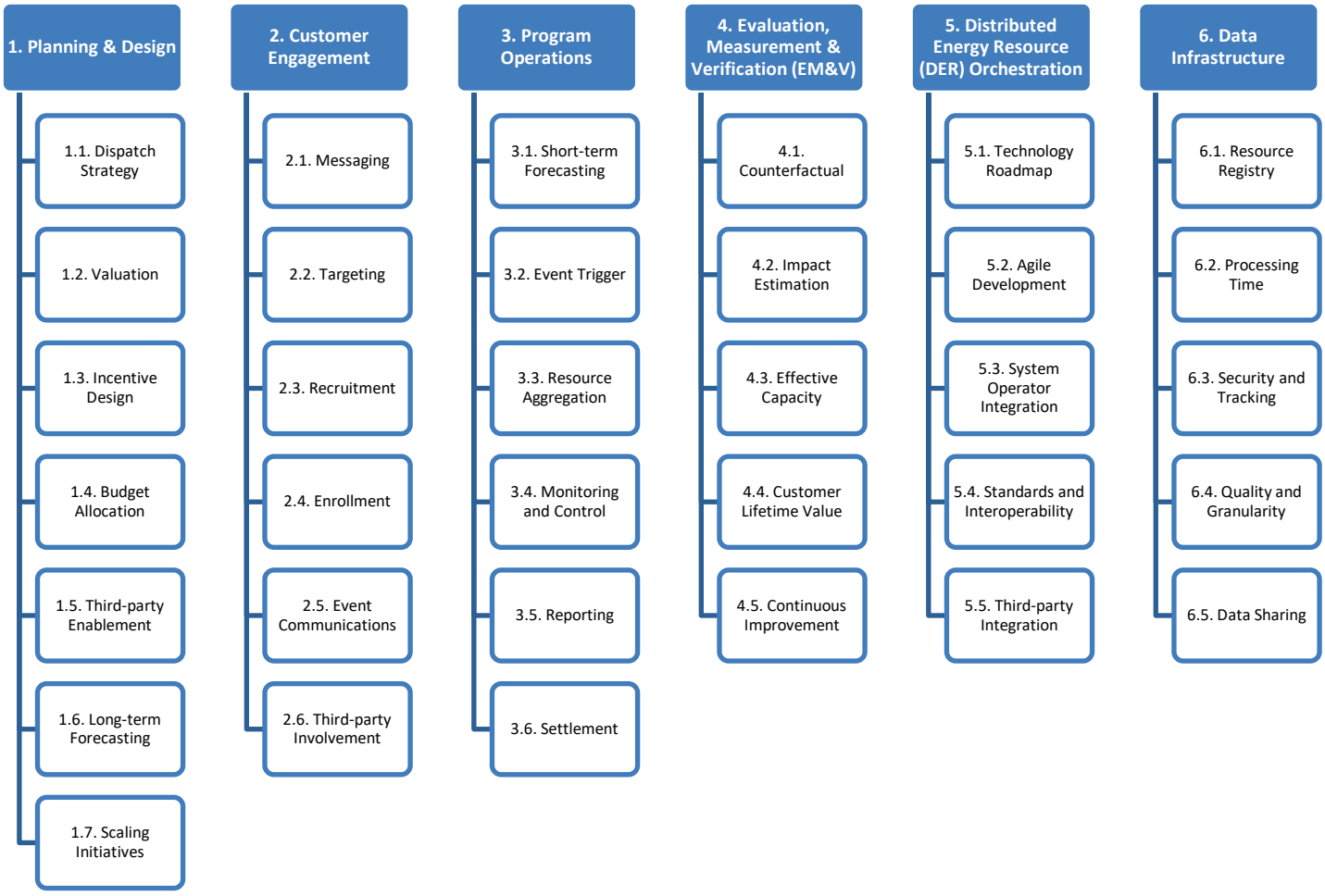


Figure 2-2. Berkeley Lab Demand Flexibility Maturity Model Categories and Capabilities

2.1.1 Planning & Design

The Planning & Design category details capabilities related to creating DF programs that are operationally effective, economically justified, and adaptable over time, including involvement of third-party DER aggregators and device makers.[6] By maturing in these capabilities, utilities can ensure that DF programs are well-positioned to support grid needs, strategic goals, and regulatory requirements with cost-effective DF resources that demonstrably offset other types of grid investments.

States and utilities can adapt the Planning & Design category based on projected grid needs, the expected role of third parties, and state objectives, priorities and valuation frameworks. Given that other categories require existing programs to assess maturity, this category can be adapted and applied on its own for new initiatives, including pilots and new programs.

[6] "Device makers" is a general term that may include everything from a smart thermostat to an electric vehicle.

2.1.2 Customer Engagement

The Customer Engagement category encompasses the strategies, tools, and processes that enable utilities to attract, onboard, and retain DF program participants in DF programs. It focuses on creating an effective customer experience that drives participation and performance, while also leveraging third parties to expand reach and enhance engagement. Importantly, this category focuses on direct customer engagement and does not cover broader marketing via social media, paid advertising, community events, and trade shows.

By maturing in these capabilities, utilities can ensure that DF programs deliver compelling value propositions, maximize participation from high-value customer segments, and foster sustained engagement. Importantly, the highest level of maturity commonly features the use of "randomized groups," which draws from best practices for causal inference in other fields.[7] Using randomized treatment and control groups has also been applied in the utilities industry to behavioral energy conservation programs and time-varying pricing (Wolfram 2018).

States and utilities can adapt the Customer Engagement category based on program objectives, customer demographics, and the anticipated role of third parties. Its applicability spans from pilots to full-scale programs, enabling tailored strategies for outreach, recruitment, and event communications that align with specific regulatory and market contexts.

2.1.3 Program Operations

The Program Operations category focuses on the capabilities necessary to effectively manage the ongoing operations of DF programs, most notably running DF events from initial short-term forecasting to dispatch and settlement. The category emphasizes operational efficiency, reliability, and performance optimization.

By maturing in these capabilities, utilities can streamline the operational aspects of DF programs, enhance coordination with grid operators, and ensure that DF resources deliver reliable and measurable benefits. Effective program operations are critical for building confidence among participants, stakeholders, and regulators, while also enabling continuous improvement through performance monitoring and reporting.

States and utilities can adapt the Program Operations category based on grid needs, resource availability, and operational frameworks. This category is particularly valuable for mature programs with established participants, but its principles can also guide pilot operations to test and refine processes before scaling.

[7] Marketing and technology developers commonly use the term "A/B testing," whereas researchers typically refer to "randomized controlled trials," including for pharmaceutical drug trials.

2.1.4 Evaluation, Measurement & Verification (EM&V)

The EM&V category focuses on the methodologies and processes used to assess the effectiveness of DF programs. This category assesses how to validate performance, optimize program design, and demonstrate the value of DF investments. By maturing in these capabilities, utilities can enhance program accountability, refine planning, and maximize customer and grid benefits. EM&V also supports regulatory compliance by ensuring transparency in program results and fostering trust among stakeholders.

States and utilities can adapt the EM&V category based on regulatory requirements, data availability, and the maturity of other interconnected capabilities, most notably within the Data Infrastructure category. EM&V capabilities apply across pilot evaluations to ongoing assessments of established DF programs, enabling continuous improvement and alignment with evolving grid and customer needs.

2.1.5 Distributed Energy Resource (DER) Orchestration

The DER Orchestration category focuses on the software integration related to operating DF resources, including on-site DERs. This includes developing the technological frameworks necessary to enable coordination among utilities, system operators, and DER aggregators. By emphasizing interoperability and scalability, this category ensures that DF resources can be effectively monitored, dispatched and controlled to meet grid needs.

By maturing in these capabilities, utilities can integrate a robust DER orchestration platform[8] that supports real-time operations, enhances grid reliability, and maximizes the value of DERs. Importantly, this category draws on common agile software development terminology, such as “sprints” and “minimum viable product.” There are many online glossaries and overviews that explain these concepts (for example, see Kaley and Krause 2023). Regardless of the specific terms used, the key for higher levels of maturity is to ensure that utilities integrate the software required to scale DF resources efficiently while maintaining adaptability to new technologies and evolving requirements.

States and utilities can adapt the DER Orchestration category based on the regulatory environment, status of existing technologies, scale required to meet grid needs, and the expected number and type of DF programs, third parties, and devices. This category is particularly useful for jurisdictions that expect to scale multiple DF programs and technologies.

2.1.6 Data Infrastructure

The Data Infrastructure category focuses on the systems and processes needed to collect, manage, and share high-quality, secure, and detailed data about DF resources and performance. It ensures that utilities and third parties can efficiently exchange validated information, enabling effective program management and decision-making. By maturing in these capabilities, utilities can build reliable, secure,

[8] “DER orchestration platform” is a general term that may include DER Management Systems, or DERMS, which are typically developed by technology providers and integrated into utility operations in an agile manner

and scalable data systems that support accurate dispatch, robust analysis, and third-party integration, forming the backbone of successful DF programs.

States and utilities can adapt the Data Infrastructure category based on the regulatory environment, status of existing technologies, scale required to meet grid needs, and the expected number and type of DF programs, third parties, and devices. While this category is particularly useful for jurisdictions that expect to scale multiple DF programs and technologies, mature Data Infrastructure capabilities are important to ensure transparency and accountability for any DF program.

2.2 Example Applications

This section includes two example maturity model applications, though there are many more potential applications. In particular, any utility, including municipal and cooperatives, can also use the DF maturity model to conduct a self-evaluation, which is how the Cybersecurity Capability Maturity Model (DOE 2022) and Smart Grid Capability Maturity Model (Carnegie Mellon 2018) are implemented.

2.2.1 Example Utility Incentive Application

This example application shows how regulators can select capabilities to use as metrics for a utility Performance Incentive Mechanism (PIM).[9] The DF maturity model starts with characterizing a utility's Dispatch Strategy capability within the Planning & Design category. It shows how a utility's maturity with respect to Dispatch Strategy evolves from a lack of alignment with system needs (inflexible resources that are rarely dispatched) to close alignment with the needs of the distribution and bulk power systems (highly flexible resources that are frequently dispatched, both locally for a subset of DF program participants and for all DF program participants).

The PIM associated with the DF portfolio could have metrics that initially focus on Planning capabilities and then evolve over time. As the utility achieves a higher level of maturity for Dispatch Strategy and other capabilities in the Planning category, the metric weighting could change to emphasize a different category (e.g., Program Operations or DER Orchestration), including third-party provider integration. The metric could also start to emphasize outcome-based performance tied to shedding, shifting and shaping load, such as the metrics that Xcel Energy reports in its Minnesota Compliance Filing (Xcel Energy 2024).

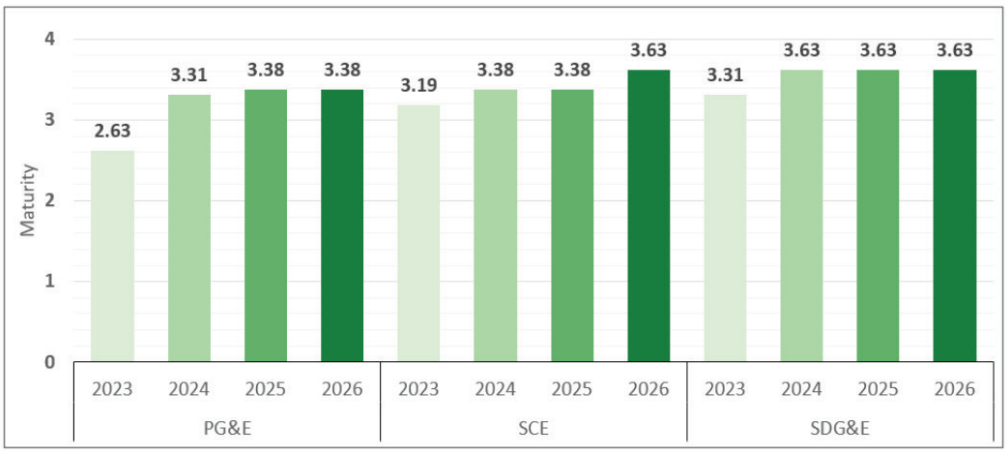
2.2.2 Example Benchmarking Application

States and utilities can also use the maturity model to benchmark capabilities. Maturity model benchmarking results are not available for DF programs, so this example draws from the California

[9] "PIMs can be implemented using a tiered approach that rewards the utility for measurable progress toward specified targets or penalizes the utility for failing to meet targets within an allotted timeframe. 'Upside-only' PIMs provide only financial benefits for achievement of specified targets, while 'downside-only' PIMs apply utility penalties when targets are not achieved. PIMs do not necessarily include both reward and penalty components, but each can be useful in driving a utility's investments and business decisions to achieve the intended outcomes" (NARUC 2023).

Wildfire Mitigation Maturity Model. Figure 2-3 benchmarks three California utilities on the current and projected maturity within the vegetation management and inspections category. By assessing the underlying capabilities within vegetation management and inspections, each utility is able to identify how to improve and can prioritize programmatic changes accordingly.

Using the DF maturity model, utilities can also benchmark capabilities and identify specific improvements. For example, three utilities may find that they have widely varying average maturity levels within the Program Operations category. Digging deeper, the utility with relatively low maturity may find that its low score for the Monitoring and Control capability is driving down its overall Program Operations score relative to the other utilities. Reviewing the capability maturity levels, the utility is able to identify how to improve the ability of program operators to review, verify, and optimize DF performance.



Source: Office of Energy Infrastructure Safety (2023)

Figure 2-3. Comparison across Utilities for Vegetation Management and Inspections Category

3. References

Carnegie Mellon (2018). Smart Grid Capability Maturity Model Compass Assessment Survey, Version 2.0. https://insights.sei.cmu.edu/library/sgmm-compass-assessment-survey-version-20

CPUC (2024). Decision Sunsetting The Investor-Owned Utilities' Demand Response Auction Mechanism Pilot Programs. https://docs.cpuc.ca.gov/PublishedDocs/Published/G000/M530/K195/530195415.PDF

Delurey, Dan (2013). Demand Response: The Road Ahead. http://www.wedgemere.com/wp-content/uploads/2013/03/Evolution-of-DR-Dialogue-Project-Report.pdf

DOE (2021). A National Roadmap for Grid-Interactive Efficient Buildings. https://gebroadmap.lbl.gov

DOE (2022). Cybersecurity Capability Maturity Model (C2M2). https://www.energy.gov/sites/default/files/2022-06/C2M2%20Version%202.1%20June%202022.pdf

DOE (2023). Pathways to Commercial Liftoff: Virtual Power Plants. https://liftoff.energy.gov/wp-content/uploads/2023/10/LIFTOFF_DOE_VVP_10062023_v4.pdf

EIA (2023). Demand response - yearly energy and demand savings. https://www.eia.gov/electricity/annual/xls/epa_10_03.xlsx

FERC (2006). Assessment of Demand Response and Advanced Metering Staff Report. https://www.ferc.gov/sites/default/files/2020-04/demand-response_4.pdf

FERC (2024). 2024 Assessment of Demand Response and Advanced Metering. https://www.ferc.gov/news-events/news/ferc-staff-issues-2024-assessment-demand-response-and-advanced-metering

Kaley, Anna and Rachel Krause (2023). Lean UX & Agile Glossary. https://www.nngroup.com/articles/agile-glossary

Murphy, Sean, Cesca Miller, Jeff Deason, Diana Dombrowski, and Portia Awuah (2024). The State of Demand Flexibility Programs and Rates. https://emp.lbl.gov/publications/state-demand-flexibility-programs-and

NARUC (2023). Demand Flexibility within a Performance-Based Regulatory Framework. https://pubs.naruc.org/pub/2A466862-1866-DAAC-99FB-E054E1C9AB13

Office of Energy Infrastructure Safety (2020). Utility Wildfire Mitigation Maturity Model. https://energysafety.ca.gov/wp-content/uploads/docs/wmp-2020/news/2.-utility-wildfire-mitigation-maturity-model-copy-correction_clean_final.pdf

Office of Energy Infrastructure Safety (2023). Decision on PG&E's 2023-2025 Wildfire Mitigation Plan. https://efiling.energysafety.ca.gov/eFiling/Getfile.aspx?fileid=56134&shareable=true

Potter, Jennifer, Elizabeth Stuart, and Peter Cappers (2018). Barriers and Opportunities to Broader Adoption of Integrated Demand Side Management at Electric Utilities: A Scoping Study. https://emp.lbl.gov/publications/barriers-and-opportunities-broader

Schellenberg, Josh, George Jiang, Candice Potter, Aimee Savage, Jeremy Smith, Andrew Hauser, Dan Lesperance, Anna-Elise Smith, and Kallie Wells (2022). Demand Response Auction Mechanism Evaluation. https://docs.cpuc.ca.gov/PublishedDocs/Efile/G000/M502/K977/502977264.PDF

Walton, Robert (2015). California's new demand response future begins. https://www.utilitydive.com/news/californias-new-demand-response-future-begins/407966

Wolfram, Catherine (2018). Smart Meters but Dumb Pricing? Not in Sacramento. https://energyathaas.wordpress.com/2013/11/18/smart-meters-but-dumb-pricing-not-in-sacramento

Xcel Energy (2024). Compliance Filing. Demand Response Compliance Docket Nos. E002/M-20-421, E002/Rp-19-368, E002/M-21-101 & E002/CI-17-401

Appendix: Demand Flexibility Maturity Model

1. Category: Planning & Design

1.1. Capability: Dispatch Strategy – Planned dispatch and criteria for when and how DF resources can be utilized

	Maturity Level				
Maturity Indicator	**1** Below expectations	**2** Meets minimum expectations	**3** Beyond minimum expectations	**4** Consistent with best practice	**5** Improvement over best practice
System needs assessment	Rudimentary understanding of current bulk power system needs	High-level understanding of current distribution and/or bulk power system needs	8,760-hour analysis of current distribution and/or bulk power system needs	8,760-hour analysis of future distribution and bulk power system needs	Thorough, sub-hourly analysis of future distribution and bulk power system needs
Notification lead time	Only with advanced notice (24+ hours) during limited time periods, such as summer weekday afternoons	Day-of notice during limited time periods, such as summer weekday afternoons	Day-of notice during more than one time period throughout the day, week and year, depending on system need	Short notice (15 minutes or less) throughout the day, week and year	Same as level 4
Resource availability	Only during grid emergency conditions (typically less than two times per year)	Primarily during grid emergency conditions (typically less than five times per year)	During grid emergency conditions, and occasionally based on economic factors (typically five or more times per year)	Dispatched regularly based on grid reliability, economic factors and/or decarbonization factors (typically 10 or more times per year)	Same as level 4
Test events			Rare test events to evaluate how DF resources respond under limited weather and system conditions	Test events to evaluate how DF resources respond under a variety of weather and system conditions	Same as level 4, with randomized groups for test events
Local dispatch				DF resources can be dispatched locally to alleviate distribution constraints or for all DF program participants to reduce system load	Same as level 4

1. Category: Planning & Design

1.2. Capability: Valuation – Process of determining the value of a DF resource from a policy perspective, based on estimated costs and benefits

	Maturity Level				
Maturity Indicator	**1** Below expectations	**2** Meets minimum expectations	**3** Beyond minimum expectations	**4** Consistent with best practice	**5** Improvement over best practice
Cost estimation	Portfolio level total	Portfolio level with primary categories (e.g., administration and incentives)	Portfolio and program level with primary categories	Program level with more granular categories (including fixed costs)	Same as level 4
Benefits estimation	Portfolio level, without using established valuation frameworks	Portfolio level using valuation frameworks that were designed for non-flexible resources (such as EE or emergency DR)	Program level based on current avoided bulk power system capacity costs (avoided distribution costs may also be included, but valuation does not align how the program is actually dispatched)	Same as level 3, with avoided capacity costs for the distribution system, accounting for location, dispatch strategy and future grid needs	Same as level 4, with accounting for effective capacity based on the forecasted magnitude and variation of load impacts
Benefit-cost Ratio (BCR) estimate		Portfolio level with limited understanding of key assumptions and value drivers within the portfolio	Program level with general understanding of key assumptions, future grid needs, and value drivers within each program	Program level by customer class, including benefits for participating customers, with sensitivity analysis to understand key assumptions, future grid needs and value drivers within each program	Same as level 4, including benefits for resiliency and decarbonization, with sensitivity analysis to identify key leading indicators that iteratively update the BCR, leading to strategic "pivots" as needed
Validation of avoided costs					Avoided costs are only used in valuation when DF resource demonstrably offsets grid investments (e.g. in an IRP), using metrics such as levelized cost of energy to directly compare to utility-scale resources

1. Category: Planning & Design

1.3. Capability: Incentive Design – Level and structure of financial compensation offered to DF participants

	Maturity Level				
Maturity Indicator	**1** Below expectations	**2** Meets minimum expectations	**3** Beyond minimum expectations	**4** Consistent with best practice	**5** Improvement over best practice
Basis for incentive levels	No clear rationale, benchmarking or alignment with valuation	Benchmarking (with publicly available data sources and/or reports)	Benchmarking (with publicly available data sources and/or reports) and general alignment with program-level BCR estimate	Same as level 3, including sensitivity analysis and alignment with program-level BCR estimate by customer class	Randomized test groups to optimize program-level BCR estimate by tailoring incentives for specific customer groups that drive value for the program
Incentive structure	Upfront incentives that do not vary based on performance or customer type	Upfront incentives that vary across residential and non-residential segments, but do not consider customer performance (such as the magnitude and reliability of load impacts)	Upfront and retention incentives that vary across residential and non-residential segments, but do not consider customer performance (such as the magnitude and reliability of load impacts)	Upfront and retention incentives that vary by customer size, location and across residential and non-residential segments, with the ability to remove customers from the program if they do not actively participate	Upfront, retention and/or performance incentives that vary by expected customer lifetime value, with the ability to remove customers from the program if they do not actively participate and enforce penalties if performance is unreliable
Accounting for other programs		Incentive eligibility accounts for participation in other DF programs or grid services to avoid double counting of the same resource	Same as level 2	Same as level 2	Incentives vary based on participation in other DF programs or grid services to maximize value of existing resources without over-compensating

1. Category: Planning & Design

1.4. Capability: Budget Allocation – Process of determining and adjusting the DF program budget

	Maturity Level				
Maturity Indicator	**1** Below expectations	**2** Meets minimum expectations	**3** Beyond minimum expectations	**4** Consistent with best practice	**5** Improvement over best practice
Basis for budget	Limited rationale, alignment with valuation or benchmarking (e.g., relative to prior years, other utilities and types of resources such as EE)	Simple benchmarks, such as budget from prior year, regardless of DF value as compared to other types of resources	Same as level 2, with high-level consideration of DF value as compared to other types of resources	Determined based on cost-effectively meeting strategic goals and regulatory requirements, with consideration of DF value as compared to other types of resources, accounting for how each DF program is dispatched and future grid needs	DF budget determined based on maximizing value (net benefits) relative to the alternatives available to address future grid needs, accounting for how each DF program is dispatched and performance incentives
Analysis of existing DF programs to inform budget		Limited	High-level (e.g., scaling of prior year cost based on projected enrollment)	In-depth, including how costs may change as the DF program evolves to meet strategic goals and regulatory requirements	In-depth, including how costs may change as the DF program evolves to maximize value (including portfolio-level investments necessary to meet goals and requirements)
Budget flexibility			Limited flexibility to reallocate program budget based on performance and changing conditions (including costs and grid needs)	Annual reviews and reallocation based on performance and changing conditions (including costs and grid needs)	Same as level 4, with quarterly reviews

1. Category: Planning & Design

1.5. Capability: Third-party Enablement – Mechanisms and processes for enabling third parties to participate in the DF program delivery

	Maturity Level				
Maturity Indicator	**1** Below expectations	**2** Meets minimum expectations	**3** Beyond minimum expectations	**4** Consistent with best practice	**5** Improvement over best practice
Third-party ability to participate	Utility-managed programs with legacy DF technologies (e.g., one-way communicating DLC devices)	Same as level 1	Same as level 1, with integration of established third-party device makers or DER aggregators	Multiple established and emerging third-party device makers or DER aggregators that enable a Level 4-5 Dispatch Strategy (see Section 1.1)	Same as level 4
Third-party opportunity		Infrequent, multi-year competitive RFP procurements for new third-party providers	Multi-year competitive RFP procurements and partnerships established third-party device makers or DER aggregators	Expanded opportunities for new third-party providers through performance-based contracts and partnerships with multiple established and emerging third-party device makers or DER aggregators that enable a Level 4-5 Dispatch Strategy (see Section 1.1)	Open opportunities for new third-party providers through flexible, performance-based contracts with any third-party device maker or DER aggregator that enables a Level 5 Dispatch Strategy (see Section 1.1)
Contracting process		Significant administrative burden for onboarding new third-party providers, enforcing penalties or terminating the agreement, regardless of performance	Same as level 2	Moderate administrative burden for onboarding new third-party providers, enforcing penalties and terminating the agreement for consistent underperformance	Minimal administrative burden for onboarding new third-party providers, enforcing penalties and terminating the agreement for unreliable performance

1. Category: Planning & Design

1.6. Capability: Long-term Forecasting – Multi-year projections of DF load impacts, based on enrollment and expected performance, for use in resource planning

	Maturity Level				
Maturity Indicator	**1** Below expectations	**2** Meets minimum expectations	**3** Beyond minimum expectations	**4** Consistent with best practice	**5** Improvement over best practice
Basis for forecast	No clear rationale, benchmarking or empirical analysis	Enrollment and event load impact forecasts determined using benchmarking (with publicly available data sources and/or reports)	Same as level 2, with high-level estimates from evaluation report for existing programs	Enrollment and hourly load impact forecasts determined using location and customer segment-specific estimates from evaluation report for existing programs, with sensitivity analysis to identify key assumptions	Same as level 4, with randomized groups to inform forecast (especially for new programs)and assessment of forecast error to evaluate uncertainty
Market potential assessment			Drives forecast assumptions and goals based on end-use adoption projections, with minimal adjustments based on evaluation report findings	Informs forecast assumptions and goals based on end-use adoption projections by customer segment, with calibration to evaluation report findings, future grid needs and expected program changes (including dispatch strategies and grid services provided)	Forecast is primarily based on the specific customer groups that are expected to drive value for the program, based on rigorous data analysis, future grid needs and expected program changes (including dispatch strategies and grid services provided)
Alignment with other programs and plans			Limited consideration of other programs and plans (includes integrated resources plans, grid modernization plans, transmission and distribution plans, electrification plans, energy efficiency program plans, etc.)	Moderate consideration of other programs and plans	Close alignment with other programs and plans

1. Category: Planning & Design

1.7. Capability: Scaling Initiatives – Planning and analysis related to scaling pilots, new program features, and new programs

	Maturity Level				
Maturity Indicator	**1** Below expectations	**2** Meets minimum expectations	**3** Beyond minimum expectations	**4** Consistent with best practice	**5** Improvement over best practice
Questions/ assumptions prior to launching initiative	No key questions or assumptions identified	Focus primarily on how a specific DF technology works with a small, unique set of customers	Focus primarily on how a specific DF technology works under current grid needs and rules	Focus on how a DF program cost-effectively scales under future grid needs	Focus on how a DF program cost-effectively scales under future grid needs, rules (e.g., FERC Order 2222), and customer expectations (e.g., a frictionless experience)
Scaling plan	No scaling plan	Limited scaling plan that applies to a narrow group of customers	Same as level 3	High-level plan for scaling the initiative to a broader group of customers, with constraints to scale as a secondary consideration for the initiative design	Detailed plan for scaling the initiative to a broader group of customers, with constraints to scale as a primary consideration for the initiative design, including selection of DF technologies, vendor partners, and third-party providers
Valuation analysis				Analyses of initiatives produce overall cost-effectiveness estimates, informing valuation, incentive design, and budget allocation for scaling plan	Analyses of initiatives produce reliable estimates of expected customer lifetime value relative to acquisition cost by customer group, informing valuation, incentive design, and budget allocation for scaling plan
Adaptive strategy				Clear articulation of goals that indicate when to move from initiative to full scale deployment	Same as level 4, with initiatives that have strategic "pivots" and contingencies tied to key leading indicators

2. Category: Customer Engagement

2.1. Capability: Messaging – Communication and tools used to convey the value proposition of DF enrollment and performance

	Maturity Level				
Maturity Indicator	**1** Below expectations	**2** Meets minimum expectations	**3** Beyond minimum expectations	**4** Consistent with best practice	**5** Improvement over best practice
Program marketing	No focused and consistent messaging	Focuses on how DF saves money	Limited messages (e.g., DF saves money and supports grid reliability)	Consistently highlights key messages (e.g., DF saves money, supports grid reliability and decarbonization, and may not cause substantial inconvenience or discomfort)	Same as level 4, with randomized groups to identify which messages are most effective for each type of customer
Participant messaging		Primarily emphasizes convenience and the ability to opt out of events	Briefly indicates the importance of frequent, automated response (if applicable), with consistent performance (including low opt-outs)	Same as level 3, with additional emphasis on the importance	Same as level 4, with randomized groups to identify which messages are most effective for each type of participant
Surveys		Verify that customers understand that DF saves money, most notably DF program participants	Occasionally verify that customers understand limited messages, most notably DF program participants	Same as level 3, with consistent frequency	Integrated into ongoing customer communications to verify that customers understand key messages assigned to randomized groups, most notably DF program participants
Customer support center				Basic understanding of DF talking points	Understands and reinforces DF messaging

2. Category: Customer Engagement

2.2. Capability: Targeting – Process of identifying and focusing efforts on high value customers

	Maturity Level				
Maturity Indicator	**1** Below expectations	**2** Meets minimum expectations	**3** Beyond minimum expectations	**4** Consistent with best practice	**5** Improvement over best practice
Basis of targeting	No customer targeting	Propensity to enroll in the DF program, regardless of customer-specific potential to deliver load impacts	Propensity to enroll in the DF program, with rudimentary screening criteria based on overall usage and customers that are known to have eligible devices	Propensity to enroll, customers that are known to have eligible devices and hourly interval data to estimate customer-specific expected DF value	Same as level 4, with sub-hourly interval data to estimate customer-specific expected DF value, using randomized groups to inform propensity model for different program options
Customer segments		Marketing personas based on predefined groups, regardless of outcomes related to DF program enrollment and usage	Same as level 2, with minor adjustments based on DF program enrollment and usage	Marketing personas based on DF program enrollment propensity, hourly usage profiles, and target segments (such as low income customers)	Customers are individually ranked based on estimates of customer-specific expected DF value, aligning with grid needs
Device adoption				Analyze hourly interval data and other customer characteristics to estimate device adoption likelihood (such as the percent chance of a customer having an EV or central AC)	Same as level 4, with the support of third-party data (e.g., DMV or dealer lists for EVs)

2. Category: Customer Engagement

2.3. Capability: Recruitment – Process and options by which the DF program engages customers to participate

	Maturity Level				
Maturity Indicator	**1** Below expectations	**2** Meets minimum expectations	**3** Beyond minimum expectations	**4** Consistent with best practice	**5** Improvement over best practice
Adaptive process	Keep recruitment as originally planned, without identifying and implementing improvements	Identify one major point of friction during recruitment (e.g., email click-through rates), then identify and implement improvements for the next program year	Same as level 2, with two or more points of friction during recruitment	Identify nearly all points of friction during recruitment, then identify and quickly implement improvements	Same as level 4, with randomized groups to test and quickly implement improvements
Enrollment options	Passively support one option for each program	Focus on supporting one primary option for each program	Passively support other enrollment options for each program, informing device makers and other incentive programs that they may also deliver recruitment materials	Streamline enrollment options for each DF program, collaborating with device makers and other incentive programs to deliver periodic co-branded recruitment materials, with default DF enrollment and a clear process to opt-out (if applicable)	Same as level 4, with consistent, co-branded recruitment materials and integration of default DF with other programs
Channels	One method to recruit customers for each program, regardless of effectiveness	Use email, phone calls and/or direct mail to recruit customers, without prioritizing the most cost-effective methods	Prioritize email recruitment, but supplement with phone calls and/or direct mail, particularly for larger non-residential accounts	Prioritize low-cost digital recruitment options, including email, text, referrals, and in-app (in partnership with device makers), but supplement recruitment with phone calls and/or direct mail, particularly for larger non-residential accounts and customers that have not engaged digitally	Same as level 4, with randomized groups to test phone calls and/or direct mail recruitment for cost-effectiveness

2. Category: Customer Engagement

2.4. Capability: Enrollment – Process and options by which the DF program fully onboards participants

	Maturity Level				
Maturity Indicator	**1** Below expectations	**2** Meets minimum expectations	**3** Beyond minimum expectations	**4** Consistent with best practice	**5** Improvement over best practice
Adaptive process	Keep enrollment process as originally planned, without identifying and implementing improvements	Identify one major point of friction during enrollment (e.g., account verification), then identify and implement improvements for the next program year	Same as level 2, with two or more points of friction during enrollment	Identify nearly all points of friction during enrollment , then identify and quickly implement improvements	Same as level 4, with randomized groups to test and quickly implement improvements
Energy audit	No upfront audit of customer energy usage	For large customers only, conduct upfront energy audit to identify DF opportunities (primarily depending on end-uses, regardless of usage)	Same as level 2, with interval data analysis to identify DF opportunities (depending on end-uses, load patterns and aggressiveness of curtailment)	Same as level 3, with streamlined audit and interval data analysis for medium and large customers	Same as level 4, with automated audit and interval data analysis for all customers
Technology installation support			High-cost support, verification, and testing available as an upfront incentive for large customers only	Support, verification, and testing available as an upfront incentive for medium to large customers, and streamlined support for high usage residential customers	Streamlined support, verification, and testing available as an upfront incentive when cost-effective based on customer-specific expected DF value
Encouraging more effective DF options				If customers are recruited into a relatively passive DF option (by default if possible), identify and implement potentially effective ways to retain them and nudge them toward more automated and effective DF	Same as level 4, with randomized groups to identify and implement most effective ways to retain customers and nudge them toward more automated and effective DF

2. Category: Customer Engagement

2.5. Capability: Event Communications – Pre- and post-event communications that inform DF participants of event timing and performance

	Maturity Level				
Maturity Indicator	**1** Below expectations	**2** Meets minimum expectations	**3** Beyond minimum expectations	**4** Consistent with best practice	**5** Improvement over best practice
Scope and messaging	No pre- or post-event communications	Basic pre-event communications via email or phone call	Same as level 2, with basic post-event communications via email or phone call	Pre- and post-event communications via email, text and/or phone call that increase the reliability of performance	Same as level 4, with randomized groups to identify most effective timing, frequency, and messaging
Post-event reports		Performance reports available in web portal two or more days later, without a focus on benefits to the customer	Performance reports available in web portal the next day, including money saved	Performance reports available in web portal the next day, including all benefits, with follow-up actions that may resolve issues impacting performance and incentives	Performance reports available in web portal within a few hours, including all benefits, pushed to customer via email and text when there are follow-up actions that may resolve issues impacting performance and incentives
Encouraging active DF			Reactive communications after a customer has not actively participated across multiple events, passively encouraging improvement	Ongoing proactive communications as a customer starts to show signs of not actively participating in events, indicating that they may be removed from the DF program	Same as level 4, with normative comparisons, offers of alternative DF program options and notices that they may be removed from the DF program and have further penalties

2. Category: Customer Engagement

2.6. Capability: Third-party Involvement – Mechanisms and tools for utilizing third parties to engage customers and manage DF resources

	Maturity Level				
Maturity Indicator	**1** Below expectations	**2** Meets minimum expectations	**3** Beyond minimum expectations	**4** Consistent with best practice	**5** Improvement over best practice
Mechanism	No third-party involvement in engaging customers	Select third-party providers focus on enrolling customers or delivering technology and provide minimal incremental value	Fixed incentives throughout service territory for multiple aggregators selected through a competitive RFP to recruit customers, onboard DF resources, and manage the interconnection process (if applicable)	Varying incentives by location for any eligible aggregator to recruit customers, onboard DF resources, and manage the interconnection process (if applicable)	Same as level 4, with a competitive marketplace (and bids to lower costs)
Tools				Utilities make maps and other tools available upon request to communicate grid needs and high value locations for DF resources	Same as level 4, with openly available maps and other tools

3. Category: Program Operations

3.1. Capability: Short-term Forecasting – Day-of or day-ahead projections of load impacts, developed by DF program to integrate DF resources into grid operations

	Maturity Level				
Maturity Indicator	**1** Below expectations	**2** Meets minimum expectations	**3** Beyond minimum expectations	**4** Consistent with best practice	**5** Improvement over best practice
Timing and granularity	No short-term forecasting of load impacts	Day-ahead forecasting of system-level average load impacts, based on previous year's evaluation or settlement results	Same as level 2, with more recent evaluation or settlement results	Day-ahead and day-of forecasting of hourly load impacts throughout the system, based on recent evaluation and settlement results, adjusted for baseline error	Day-ahead and day-of forecasting of hourly load impacts throughout the system, down to the feeder level, based on recent evaluation results (using Level 4-5 EM&V methods in Section 4) and varying control strategies
Adjusting for recent data and forecasts			Minor adjustments based on recent trends, such as enrollment	Potentially significant adjustments based on weather forecasts and recent data related to load patterns and enrollment for each DF program participant	Potentially significant adjustments based on weather forecasts and near real-time data related to device connectivity, opt outs, overrides, load patterns, and enrollment for each DF program participant

3. Category: Program Operations

3.2. Capability: Event Trigger – Criteria for dispatching and prioritizing DF resources

	Maturity Level				
Maturity Indicator	**1** Below expectations	**2** Meets minimum expectations	**3** Beyond minimum expectations	**4** Consistent with best practice	**5** Improvement over best practice
Criteria for DF dispatch	Bulk power system day-ahead emergency alerts with advanced notice (24+ hours), as long as resource is available at that time	Same as level 1, with day-of emergency alerts (notice of over 15 minutes)	Same as level 2, with day-ahead price thresholds	Transmission/distribution constraints and bulk power system day-ahead price thresholds or day-of emergency alerts (with notice of 15 minutes or less when necessary)	Same as level 4, with real-time price thresholds or emergency alerts
Coordinating with grid operators				Limited coordination between distribution and bulk power system operators to ensure that resources are triggered in priority order	Same as level 4, with close coordination
Other grid services					DF is dispatched to provide other grid services (e.g., frequency response) or support grid decarbonization based on carbon intensity signals or renewable generation needs

3. Category: Program Operations

3.3. Capability: Resource Aggregation – Process of combining, prioritizing, and ordering DF resources for dispatch

	Maturity Level				
Maturity Indicator	**1** Below expectations	**2** Meets minimum expectations	**3** Beyond minimum expectations	**4** Consistent with best practice	**5** Improvement over best practice
Objective	Maximize the initial load reduction for the bulk power system	Same as level 1	Maximize the initial load reduction, accounting for the timing of bulk power system needs and forecasted average load impacts	Maximize value to the grid, accounting for the timing and location of grid needs, expected event duration, and forecasted hourly load impacts, including increases in load before and after the event	Same as level 4, with accounting for customer fatigue and forecasted sub-hourly load impacts, including increases in load before and after the event
Scale				Sufficient scale to monitor and accurately measure load impacts for settlement and evaluation purposes	Same as level 4, with randomized groups when a portion of the resource can be withheld
Accounting for other programs					DF aggregations are prioritized for dispatch across markets (if applicable) and levels of the system, based on maximizing value to the grid, accounting for when there are multiple DERs in the same location

3. Category: Program Operations

3.4. Capability: Monitoring and Control – Ability of program operators to review, verify, and optimize DF performance

	Maturity Level				
Maturity Indicator	**1** Below expectations	**2** Meets minimum expectations	**3** Beyond minimum expectations	**4** Consistent with best practice	**5** Improvement over best practice
Program operator role	Dispatch DF resources and review DF performance at end of season	Same as level 1, with review of DF performance days later	Same as level 2, with review of DF performance the next day	Optimize dispatch based on grid needs and review DF performance shortly after events	Same as level 4, with DF performance monitoring in near real-time
Dispatch verification	One-way dispatch signal without verification that device received signal	Same as level 1, with ad hoc verification based on interval data two or more days later	One-way or two-way dispatch signal with next-day verification based on interval data or device-level data	Two-way dispatch signal with verification that has a delay of 5 minutes or longer	Two-way dispatch signal with near real-time verification and connectivity checks leading up to and throughout DF events

3. Category: Program Operations

3.5. Capability: Reporting – Post-event reports for operators of grid, DF program and DERs to identify performance issues and provide initial load impact estimates

	Maturity Level				
Maturity Indicator	**1** Below expectations	**2** Meets minimum expectations	**3** Beyond minimum expectations	**4** Consistent with best practice	**5** Improvement over best practice
Level of detail	No event reporting	Basic event reports that focus on average load impact during event period using a settlement baseline	Same as level 2, with hourly load impacts	In-depth event reports that detail performance issues and hourly load impacts during event period using a settlement baseline and evaluation counterfactual	Same as level 4, with assessment of customer fatigue and sub-hourly load impacts
Report availability		Two or more days later and not shared with other parties	Next day and shared with other parties upon request	Shortly after events and automatically shared with other parties, particularly grid operators and those who may help resolve performance issues	Same as level 4, with notifications during events and follow-ups to resolve issues

3. Category: Program Operations

3.6. Capability: Settlement – Process of determining financial compensation for DF participants

	Maturity Level				
Maturity Indicator	**1** Below expectations	**2** Meets minimum expectations	**3** Beyond minimum expectations	**4** Consistent with best practice	**5** Improvement over best practice
Baseline	Same for all customers and has not been validated for accuracy, including reliance on simplifying assumptions (e.g., an EV would have been plugged in and charging throughout a DF event)	Same for all customers and has been validated for accuracy when applied to large aggregations of DF resources	Varies by customer type and has been validated for accuracy when applied to aggregations of DF resources by customer type	Same as level 3, with quantification of over- and under-payment	Same as level 4, with use of comparable groups of non-participating customers to further assess accuracy
Processing time	Over 96 hours after event, primarily due to accessing and validating data	Up to 96 hours after event	Up to 8 hours after event	Up to 2 hours after event, with automated data validation	Up to 30 minutes after event, with continuous data validation and quick access
Basis for capacity payments (if applicable)	Unvalidated load impact estimates	Validated load impact estimates for a single event	Validated load impact estimates across multiple events	Validated load impact estimates that are consistent at the hourly level throughout events (e.g., capacity payments are not based on a single event hour estimate)	Same as level 4, with consistency at the sub-hourly level

4. Category: EM&V

4.1. Capability: Counterfactual – Estimate of what load would have been in the absence of a DF event, which informs load impact estimates

	Maturity Level				
Maturity Indicator	**1** Below expectations	**2** Meets minimum expectations	**3** Beyond minimum expectations	**4** Consistent with best practice	**5** Improvement over best practice
Establishing counter-factual	No control or comparison group, leading to a reliance on simplifying assumptions (e.g., an EV would have been plugged in and charging throughout a DF event)	Comparison group developed without statistical load matching	Matched control group with moderate differences in peak load during non-event days with the highest system load	Matched control group without statistically significant differences in peak load during non-event days with the highest system load	Randomized control group without statistically significant differences in peak load during non-event days with the highest system load
Role of settlement baseline	Relied on for evaluation counterfactual	Option for evaluation counterfactual due to inaccuracy of comparison group	Not used for evaluation counterfactual due to relative accuracy of control group	Same as level 3	Same as level 3

4. Category: EM&V

4.2. Capability: Impact Estimation – Process of evaluating and quantifying the DF program load reductions and shifting

	Maturity Level				
Maturity Indicator	**1** Below expectations	**2** Meets minimum expectations	**3** Beyond minimum expectations	**4** Consistent with best practice	**5** Improvement over best practice
Calculation	Difference between event usage and assumed counterfactual or settlement baseline (if applicable) for all customers	Difference between event usage and comparison group for all customers	Estimated load impacts using matched control group and difference-in-differences	Same as level 3, with a regression model that estimates confidence bands	Same as level 4, with a randomized control group
Granularity	Single average event impact estimate based on limited events during rare grid emergency conditions	Average impacts by event day during limited time periods, such as summer weekday afternoons	Hourly impacts by event day for five or more events in a year across more than one time period in the day, week or year (including test events)	Hourly impacts by event day for ten or more events in a year under a variety of weather and system conditions (including test events)	Same as level 4, with sub-hourly impacts by event day
Load shifting analysis			Hourly pre- and post-event load shifting estimates to evaluate the full DF impact	Same as level 3, with analysis of the net benefits of pre- and post-event load shifting	Same as level 4, with sub-hourly pre- and post-event load shifting estimates
Customer segments				Load impact estimates by dispatch group to account for the varying number of customers that participate in each event (primarily due to local dispatches)	Same as level 4, with load impact estimates by key customer characteristics to identify high/low value segments

4. Category: EM&V

4.3. Capability: Effective Capacity – Estimate of DF resource capacity adjusted for variation of average load impacts, contribution to grid reliability, and other factors

	Maturity Level				
Maturity Indicator	**1** Below expectations	**2** Meets minimum expectations	**3** Beyond minimum expectations	**4** Consistent with best practice	**5** Improvement over best practice
Basis for DF capacity estimate	Single average event impact from limited events during rare grid emergency conditions	Average impact across event days during limited time periods, such as summer weekday afternoons	Relationship between temperature and average impact per customer for each event day to estimate capacity under varying weather conditions and enrollment scenarios	Same as level 3, with hourly impacts per customer for each dispatch group	Relationship between temperature and sub-hourly impact per customer for each event day to estimate capacity under varying weather conditions, times of day, event durations, grid service type and enrollment scenarios for each dispatch group
Approach to adjusting capacity			Based on system planning input and models, adjust capacity based on the magnitude and variation of average load impacts	Same as level 3, with hourly load impacts and consistent use of adjusted capacity metrics such as Effective Load Carrying Capability (ELCC), which reflects contribution to meeting grid reliability needs	Same as level 4, with accounting for variation of sub-hourly load impacts and availability based on how well DF resource has aligned with grid needs
Accounting for other programs			Simplifying assumptions to adjust for overlapping capacity of other DF programs	Identify dual-enrolled customers to adjust for overlapping capacity of other DF programs	Same as level 4, with accounting for which program takes precedence for planners and system operators
Increases in load					Consider capacity value of increases in load when needed, such as during renewable curtailments

4. Category: EM&V

4.4. Capability: Customer Lifetime Value – Metric that represents the estimated total net benefits expected from a customer throughout DF program enrollment

	Maturity Level				
Maturity Indicator	**1** Below expectations	**2** Meets minimum expectations	**3** Beyond minimum expectations	**4** Consistent with best practice	**5** Improvement over best practice
Value estimation	No estimation of DF benefits	Evaluate DF program-level cost-effectiveness based on load impacts and minor updates to planning assumptions based on evaluation results	Same as level 2, with further updates to planning assumptions based on evaluation results	Evaluate Customer Lifetime Value (CLV) based on the present value of net benefits, using effective capacity to estimate avoided costs, and ongoing retention costs, adjusting for historical attrition	Evaluate CLV based on the present value of net benefits by customer group, using effective capacity to estimate avoided costs, and ongoing retention costs (including excess payment due to baseline error), adjusting for attrition using a rigorous customer retention forecast
Considering customer acquisition cost				Evaluate the average Customer Acquisition Cost (CAC) and calculate CLV-to-CAC ratio to assess the value of each participant as the DF program scales	Same as level 4, with analysis by customer group to assess the value of each type of participant as the DF program scales
Applying CLV				Evaluate the impact of CLV and CAC on overall DF program cost-effectiveness based on various enrollment scenarios	Calculate the break-even point of the required number of each type of participant to cover ongoing costs that support program operations and portfolio-level investments, and compare to achievable number of potential participants

4. Category: EM&V

4.5. Capability: Continuous Improvement – Ongoing effort to identify changes that increase DF program value based on data, metrics and EM&V results

	Maturity Level				
Maturity Indicator	**1** Below expectations	**2** Meets minimum expectations	**3** Beyond minimum expectations	**4** Consistent with best practice	**5** Improvement over best practice
Nature of process	Unstructured and reactive, focused on mitigating damage as problems arise, without data-driven insights from prior program cycle or other utilities	Primarily reactive, focused on mitigating damage as problems arise, with limited data-driven insights from prior program cycle or other utilities	Reactive and proactive, focused on identifying opportunities with data-driven insights from prior program cycle or other utilities	Primarily proactive, focused on identifying opportunities with data-driven insights from prior program cycle and other utilities	Same as level 4, with focus on identifying opportunities to maximize value (net benefits)
Monitoring changes made		Basic plan for monitoring high-level metrics, but no effort to identify further improvement opportunities	Basic plan for monitoring key metrics and identifying further improvement opportunities	Same as level 3, with a detailed plan	Same as level 4, with randomized groups, monitoring granular metrics and identifying further improvement opportunities tied to results of testing
Engaging other parties				Utility leadership, grid operators and stakeholders are moderately engaged in the continuous improvement process	Same as level 4, with more engagement and clearly defined metrics that are broadly understood

5. Category: DER Orchestration

5.1. Capability: Technology Roadmap – Strategic plan that outlines the steps and technology features required to scale a platform that meets DER orchestration needs

	Maturity Level				
Maturity Indicator	**1** Below expectations	**2** Meets minimum expectations	**3** Beyond minimum expectations	**4** Consistent with best practice	**5** Improvement over best practice
Nature of roadmap	No roadmap for DER orchestration platform	Reactive and focused on up to two DF programs	Focused on current grid needs and up to four DF programs	Prioritized features that align with near-to-long term planning and grid needs	Prioritized features that align with near-to-long term grid needs and value, with more detail for near-term features
Frequency of updates		Roadmap is static and rarely changes	Roadmap changes annually as new requirements, DF resources, and devices are identified, with over one year lead times to deliver new features	Roadmap changes semi-annually, integrating open standards such as OpenADR and IEEE 2030.5, with less than one year lead times to deliver new features	Same as level 4, with quarterly changes and less than 6-month lead times to deliver new features
Integrating other DF programs				Defined plan to iteratively integrate nearly all DF programs onto one DER orchestration platform to allow for multi-DER management and cross-DER optimization	Same as level 4, with plan to integrate all DF programs

5. Category: DER Orchestration

5.2. Capability: Agile Development – Flexible and iterative approach for deploying a DER orchestration platform that adapts to evolving requirements while mitigating risks

	Maturity Level				
Maturity Indicator	**1** Below expectations	**2** Meets minimum expectations	**3** Beyond minimum expectations	**4** Consistent with best practice	**5** Improvement over best practice
Planning approach	No plan for a DER orchestration platform	Long-term planning using a waterfall development process, with limited agreement on highest priority near-term needs	Asynchronous sprint planning with a vaguely defined minimum viable product (MVP) and limited agreement on highest priority needs	Joint sprint planning with a clearly defined MVP, using diagrams and mockups to capture and agree on highest priority needs	Same as level 4
Iteration process		Infrequent demos of new features to utility users before release	Two-week sprints with sporadic, asynchronous demos of new features (when available) so that utility users can reactively validate DER orchestration enhancements and their ability to meet DF goals and regulatory requirements	Same as level 3, with live demos of new features (when available) to enable proactive validation	Same as level 4, with live demos of distinct features that can be delivered by the end of each sprint

5. Category: DER Orchestration

5.3. Capability: System Operator Integration – Ability of distribution and bulk power system operators[10] to monitor and control DF resources via a DER orchestration platform

	Maturity Level				
Maturity Indicator	**1** Below expectations	**2** Meets minimum expectations	**3** Beyond minimum expectations	**4** Consistent with best practice	**5** Improvement over best practice
Role of distribution system operator	No interaction with a DER orchestration platform	Provide input on when to dispatch all available DF resources in response to a grid emergency	Provide input on how to optimize DF dispatch based on grid needs, and receive DF performance reports the next day	Control DF resources via Advanced Distribution Management System (ADMS) integration, optimize dispatch based on grid needs, and receive DF performance reports shortly after events	Same as level 4, with ability to monitor DF performance in near real-time
Role of bulk power system operator, including ISOs and RTOs (if appliable)	No interaction with a DER orchestration platform	Provide input on when to dispatch all available DF resources in response to a grid emergency	Provide input on how to optimize DF dispatch based on grid needs, and receive DF performance reports the next day	Control DF resources via operational systems integration, optimize dispatch based on grid needs, and receive DF performance reports shortly after events	Same as level 4, with ability to monitor DF performance in near real-time

[10] The roles that operators of the distribution system and bulk power system (generation and transmission) vary by jurisdiction. In some cases, one DF program may be operated at both the distribution and bulk power levels, in which case close coordination between operators is required. This coordination and integration between distribution and bulk power system operators is outside the scope of the maturity model.

5. Category: DER Orchestration

5.4. Capability: Standards and Interoperability – Scalable rules and guidelines for dispatching, managing, and exchanging data with DF resources, including devices and controls

	Maturity Level				
Maturity Indicator	**1** Below expectations	**2** Meets minimum expectations	**3** Beyond minimum expectations	**4** Consistent with best practice	**5** Improvement over best practice
Approach	Custom development for one DF program	Use available open standards and interoperability for data, devices, and controls	Same as level 2	Prioritize open standards and interoperability for data, devices, and controls to have one system that can cost-effectively scale by integrating with a wide range of DER aggregators and technologies	Same as level 4, with co-development of open standards
Proprietary standards and custom APIs		Multiple required to integrate more than two DF programs	Multiple required to integrate all DF programs	Used as a secondary option when necessary to integrate all DF programs	Same as level 4
Testing new integrations		Reactive testing after the technology onboarding process (including alignment with any applicable open standards such as OpenADR and IEEE 2030.5)	Same as level 2, with spot testing during the technology onboarding process	Same as level 3, with thorough testing during the technology onboarding process	Thorough, proactive testing and verification of consistent performance and alignment with any applicable open standards such as OpenADR and IEEE 2030.5, including improvements to standards that are co-developed

5. Category: DER Orchestration

5.5. Capability: Third-party Integration – Process of onboarding new device and aggregations and validating DF resources

	Maturity Level				
Maturity Indicator	**1** Below expectations	**2** Meets minimum expectations	**3** Beyond minimum expectations	**4** Consistent with best practice	**5** Improvement over best practice
Onboarding new devices and aggregators	No onboarding process	Slow and manual process with significant administrative burden for all parties involved, especially when transferring DF resources between aggregators or programs	Semi-automated process with moderate administrative burden for all parties involved	Semi-automated, user-friendly process with limited administrative burden	Automated, user-friendly process with minimal administrative burden
Validating DF resources		Limited, reactive validation of third-party DF commitments that only identifies major issues after a prolonged period of underperformance	Same as level 2, with validation shortly after significant underperformance issues arise	Same as level 3, with proactive validation of third-party DF commitments using load data	Same as level 4, with ongoing performance-based adjustments to third-party DF commitments as events are called, including test events

6. Category: Data Infrastructure

6.1. Capability: Resource Registry – Centralized database that tracks and manages information about attributes and dispatches of DF resources, including on-site DERs

	Maturity Level				
Maturity Indicator	**1** Below expectations	**2** Meets minimum expectations	**3** Beyond minimum expectations	**4** Consistent with best practice	**5** Improvement over best practice
Availability of validated DF resource attributes	Not validated or readily available	DF resources are tied to a utility account and meter ID, but information could be outdated and requires at least 4 hours of manual effort to manage and report on each time	Same as level 2, with up to 4 hours of manual effort to manage and report on each time	DF resources are tied to a frequently validated utility account and meter ID that are readily available with limited manual effort, but other resource information could be outdated and requires up to 4 hours of manual effort to manage and report on each time	Central registry of DF resources with validated information on location, technology, utility account, meter ID, eligibility checks, availability, and capacity, all of which are readily available with limited manual effort
Availability of DF dispatch information			Limited information on resource availability, dispatches, and response available	Resource availability, dispatches, and response are tracked at the utility account and meter ID level, including critical information for certain technology types (such as consumption, discharge, and state of charge for EVs and batteries)	Same as level 4

6. Category: Data Infrastructure

6.2. Capability: Processing Time – Duration it takes for a system to process DF resource and dispatch data from its input stage to validated output

	Maturity Level				
Maturity Indicator	**1** Below expectations	**2** Meets minimum expectations	**3** Beyond minimum expectations	**4** Consistent with best practice	**5** Improvement over best practice
DF resource data	No reliable data within 48 hours	Disparate datasets that require up to 48 hours to access and validate for enrollment reporting and processing purposes	Same as level 2, with up to four hours to access and validate	Same as level 3, with up to one hour to access and validate	Central registry that is continuously validated, updated, and readily available to access for enrollment reporting and processing purposes
DF dispatch data	No reliable data within 48 hours	Requires up multiple days to access and validate for settlement purposes and event reporting that is available to participants, aggregators, system operators and other stakeholders	Same as level 2, with up to one day to access and validate	Same as level 3, with up to two hours to access and validate	DF dispatch data is continuously validated, updated, and available within 30 minutes to access for settlement purposes and all event reporting
System operator needs					Based on system operator requirements, select DF dispatch data is available in near real-time to closely monitor and control resources during events

6. Category: Data Infrastructure

6.3. Capability: Security and Tracking – Processes and practices for securely sharing and storing data and tracking how source data has been validated and managed

	Maturity Level				
Maturity Indicator	**1** Below expectations	**2** Meets minimum expectations	**3** Beyond minimum expectations	**4** Consistent with best practice	**5** Improvement over best practice
Sharing and storage security	Data is shared manually between parties, including via email and stored locally on employee computers	Sensitive customer information is only shared and stored using secure, approved means, but aggregated reports may be shared via email and stored locally	Same as level 2	All data are only shared and stored using secure, approved means, including aggregated reports	Same as level 4
Tracking customer data	No reliable customer data tracking	All parties track sensitive customer information and can permanently delete the data upon request, but it requires a manual process that takes up to a week	Same as level 2, with a manual process that takes up to a day	Automated tracking to revoke access to customer data upon request within 1 hour, ensuring that all parties no longer have access to sensitive information	Same as level 4, with ability to immediately revoke access
Tracking original source data			Limited tracking and inconsistent documentation on how source data has been validated and managed in subsequent datasets	Detailed tracking and consistent documentation on how source data has been validated and managed in subsequent datasets	Same as level 4, with continuously updated documentation, including data that originate from third parties

6. Category: Data Infrastructure

6.4. Capability: Quality and Granularity – Overall reliability and level of detail of the information captured within a dataset

	Maturity Level				
Maturity Indicator	**1** Below expectations	**2** Meets minimum expectations	**3** Beyond minimum expectations	**4** Consistent with best practice	**5** Improvement over best practice
Data to inform load impact estimation	No validated data	Limited (e.g., device-level runtime data, kWh readings that are not revenue-quality or available in 1-hour intervals)	Validated data based on revenue-quality kWh readings is available in intervals of up to 1 hour, with corroboration by device-level data when available	Same as level 3, with intervals of up to 15 minutes	Same as level 4, with intervals of up to 5 minutes
Time and location of dispatch signals		Based on DF resource participation start and end dates	Same as level 2, with limited raw data on dispatch signals available at the individual customer level	Validated data on the time (to the minute) and location of every dispatch signal is available at the individual customer level for all DF resources, including those dispatched by third parties	Same as level 4, with two-way verification of every dispatch signal
Checks of data quality and availability		Ad hoc when major errors or gaps arise, with workarounds to temporarily resolve the issue	Same as level 2, with initial checks before going live	Weekly (or more frequent) monitoring of data quality and availability, with manual notifications to DF program managers, customers, and third parties (if applicable) when data errors or gaps arise	Same as level 4, with continuous monitoring of data quality and availability and automatic notifications
Third-party data providers, including device makers and aggregators			Limited reactive communication to clarify data units and interpretations, fix data structures that change and update systems accordingly	Same as level 3, with more frequent communication to resolve data issues quickly	Same as level 4, with proactive communication to identify when data structures may change and update systems accordingly

6. Category: Data Infrastructure

6.5. Capability: Data Sharing – Transfer and consistency of data exchange to and from third parties

	Maturity Level				
Maturity Indicator	**1** Below expectations	**2** Meets minimum expectations	**3** Beyond minimum expectations	**4** Consistent with best practice	**5** Improvement over best practice
Data transfers	Reliance on manual transfers for data sharing to and from third parties	Certain data transfers to and from third parties are well documented, which includes identification of the source data and all subsequent data management steps	Same as level 2, with most data transfers being well documented	Same as level 3, with most data transfers being automated, streamlining routine data-related processes (e.g., accessing interval data after a third party receives customer consent, verifying a customer account and eligibility when a third party enrolls participants)	Nearly all data transfers to and from third parties are automated and well documented
Data and reporting between the utility and third parties	Frequently inconsistent due to unclear processes, human error, and a lack of validated data that is accurately interpreted	Inconsistent for data transfers that are not well documented	Same as level 2	Same as level 3	Rarely inconsistent
Burden for all parties involved with data sharing	Overwhelming	Unsustainable	Significant	Manageable	Limited