

# Provider data strategy for sustainable data quality and accuracy





## Introduction

There has been significant scrutiny on provider directory accuracy in the past five years, with regulations at the federal and state levels already enacted. This started with a Government Accountability Office (GAO) audit of 100+ Medicare Advantage health plans that revealed an average provider directory accuracy score of about 50%. An accurate provider directory is critical for plan members to find and receive care. It could also have a fiscal impact on members and providers in case of data inaccuracies about the provider network status.

Much of the focus of regulatory measures has been to hold payers accountable for poor data quality that has not provided the desired results. The majority of the provider data is derived from sources that are intrinsically of substandard quality. The data then undergoes further deterioration in quality due to poor data management practices and legacy foundations

in the payer ecosystem. Meeting new regulations has required heavy investments by both payers and providers to upgrade their data infrastructure and business processes. As per CAQH estimate, the provider directory data maintenance alone accounts for over \$2.3B in annual administrative expenses for payers and providers, which has increased in the last two years to comply with the new mandates.


A more recent development is CMS's public notice to create a centralized data hub. The API-enabled National Directory of Healthcare Providers and Services (NDH) will aggregate provider directory data across all providers to a centralized data platform. While this will ease the burden on payers, providers will be challenged to comply with these measures in the short term. With many other compliance reports such as the National Plan and Provider Enumeration System (NPPES) and the Provider Enrollment Chain and Ownership System (PECOS), American Hospital Association (AHA) it also highlighted concerns about the duplication of effort and the lack of readiness for FHIR-based API integrations for such an initiative.


Provider data is much more diverse, with provider directory attributes just a small subset. As noted


in the whitepaper '[Provider lifecycle management \(PLM\) for a digital enterprise](#)', payers and providers collectively face challenges with the lack of standard data and processes to exchange a universe of provider data between two entities that cause data quality issues. This results in inefficiencies, high administrative costs and data quality issues such as provider directory inaccuracy, claim and authorization fallouts, delay in provider onboarding etc. It is, therefore, essential to address the gaps with a holistic provider data solution rather than focusing on the provider directory alone.


The current state payer and provider ecosystems are not built to manage future expected outcomes due to their legacy infrastructure and outdated architecture. Health plans must invest in their provider lifecycle foundations to prepare for a digitally integrated ecosystem that may include a centralized provider directory as another data source and, preferably, a holistic data hub for the provider data universe. A calibrated provider data approach can set the payers apart from their competitors.

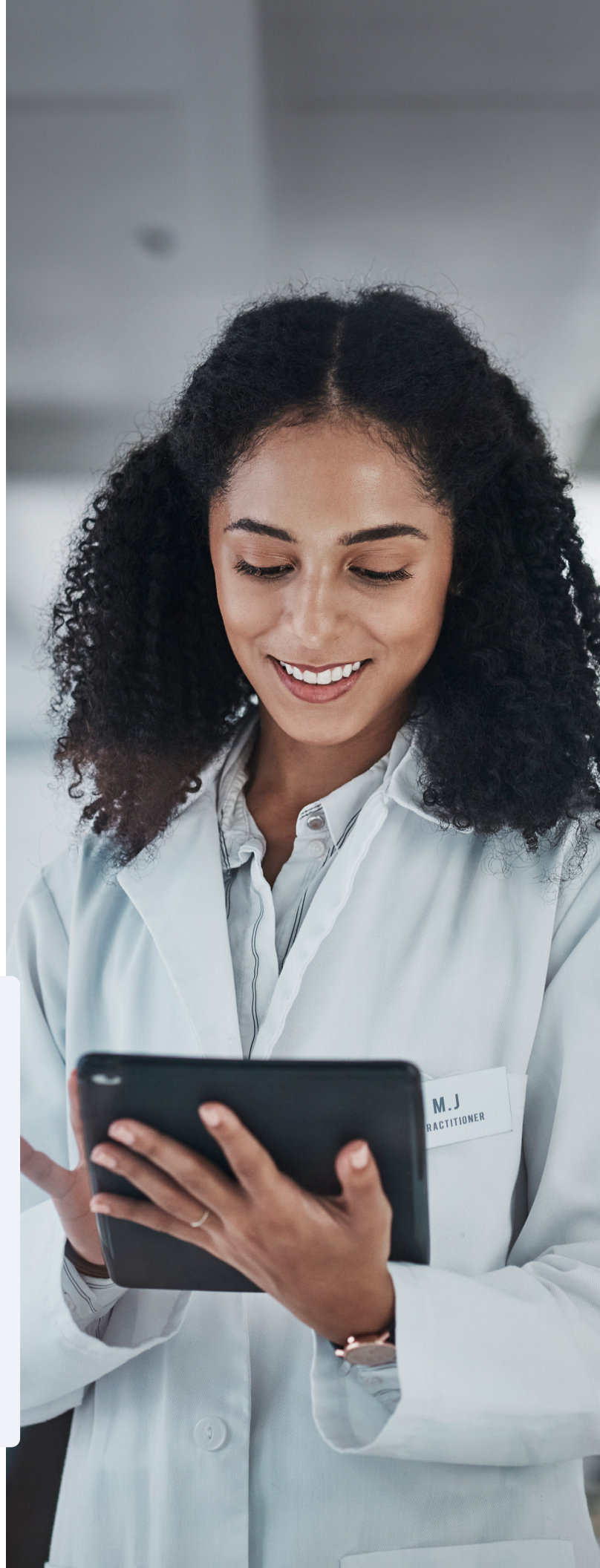
At HCLTech, we view the provider data in the context of end-to-end provider lifecycle management processes. Our solution and recommendation are to build a holistic data strategy that provides a sustainable level of data quality using a four-pronged approach:

 Scalable and agile data foundation to enable provider 360 view.

 Data diagnostics to detect and cleanse insufficient data.

 Master data management and governance for sustainable data quality.

 An intelligence engine that drives proactive insights and self-cleansing.

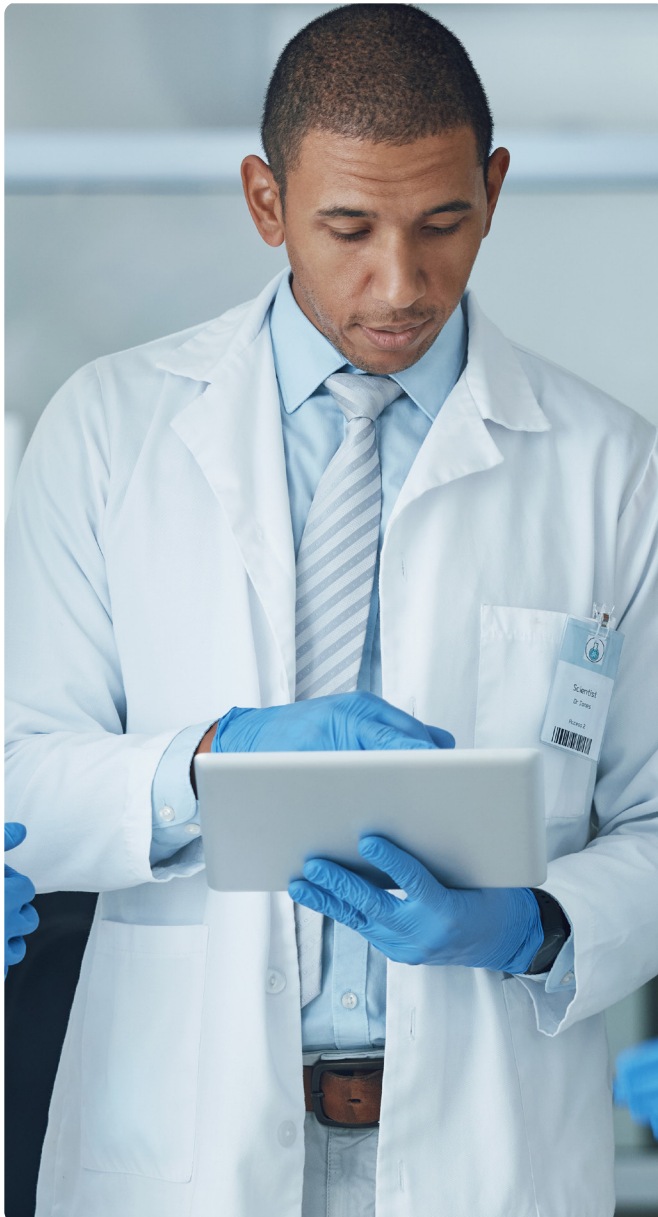


# Key industry changes

Below are the vital drivers that require health plans to build a holistic provider data strategy

## Mainstreaming value-based care

The success of value-based care initiatives hinges on the accurate and timely data sharing between health plans and providers. The provider data available to health plans drives provider engagement and maximizes the gains from value-based care initiatives. However, poor data quality and accuracy have created a need to reimagine the provider data lifecycle from origin to consumption.



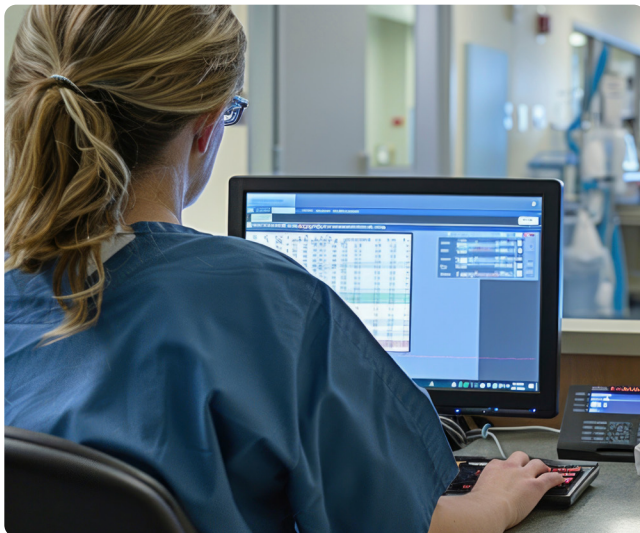
## Adhering to regulatory compliance

In 2017, GAO first audited and reported a dismally low provider directory accuracy of approximately 50% across the health plans servicing Medicare Advantage plans. Despite many federal and state regulations (including the No Surprise Act, California SB 137 etc.), the directory accuracy has not improved much over the last five years, even with the threat of penalties and sanctions. The current set of regulations is mainly focused on provider directory datasets. CMS recently released an RFI to invite public comments for a National Directory of Healthcare Providers and Services (NDH) that could function as a centralized data hub for all healthcare directories and digital contact information. While this will help address directory accuracy issues, it will limit the extent to which payers and providers can improve the overall provider data quality and accuracy. Provider directory attribute, although impactful, forms an exceedingly small subset of the provider data universe that is exchanged with various ecosystem players. As a result, providers will continue to share data in multiple custom formats which will continue the gaps associated with the source data in provider systems for other business processes.

## Real-time standard process for data exchange using data interoperability



Federal regulations have initiated real-time data exchange within and across the healthcare ecosystem using FHIR-based interoperability in multiple areas including clinical data sharing, prior authorization and others. The data standardization and interoperability requirements will immensely benefit provider data exchange with health plans. The vision for the national provider directory will facilitate provider directory data exchange from providers to CMS and CMS to payers, but the scope needs to expand further to include the holistic data exchange.



## Platform consolidation/modernization



Mergers and acquisitions in the healthcare industry result in disparate provider data platforms, creating a need to consolidate regional systems into enterprise-level platforms. Many health plans use legacy provider data platforms that lack the MDM, workflow management and analytics capabilities required to support new regulations and initiatives like value-based care. This is driving health plans to adopt cloud-enabled next-generation data platforms that support MDM, workflow management and analytics in their DNA.

## Enhancing provider experience through effective engagement



The payer-provider collaboration requires addressing the lack of trust and transparency that have existed historically between the two entities. Provider lifecycle management processes (contracting, onboarding, data maintenance and servicing) have traditionally been viewed as a back-office function lacking focus on providers' digital engagement, resulting in poor onboarding experience, miscommunications and data challenges. Better management of provider data through digital engagement and servicing that makes the process more transparent will improve providers' experience with the health plan and help build trust with each other.



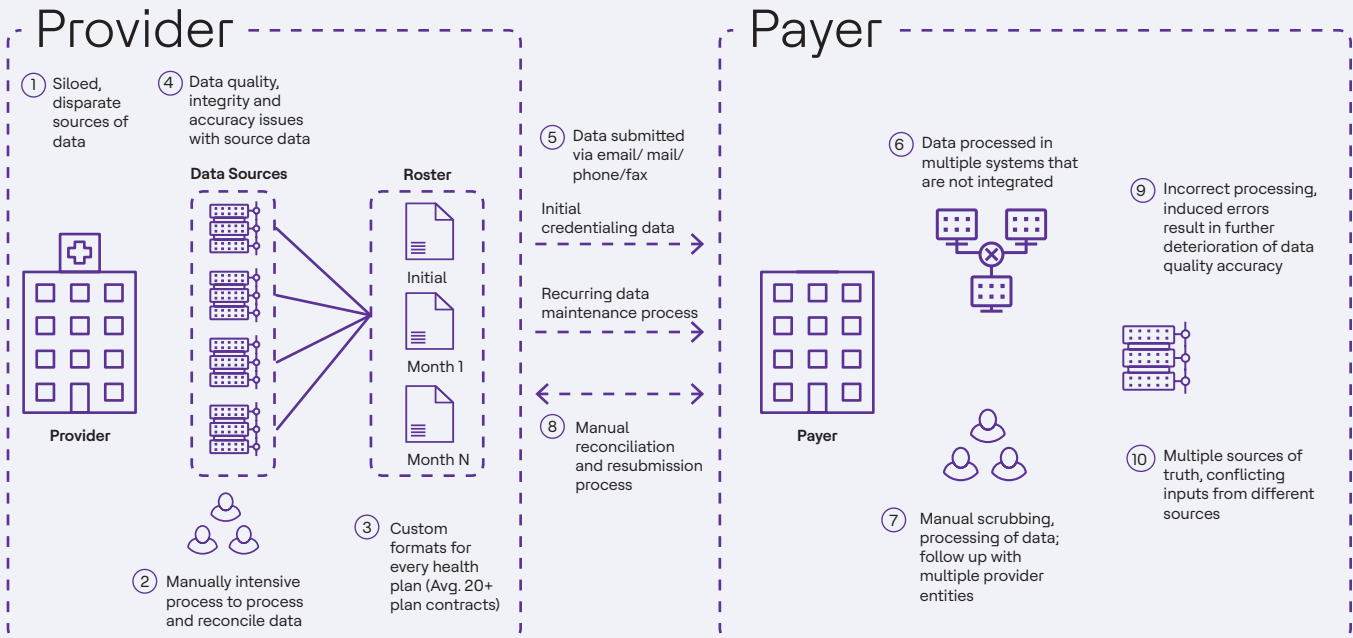
# Industry challenges and impacts

## Current state provider-payer data exchange process

When payers and providers engage in complex business transactions manually or through legacy technologies, they produce data that is often of poor quality and inaccurate, making it difficult to manage.

Below is a high-level data exchange process between payers and providers:

Complex business transactions when performed manually or using legacy technologies, generate data with unmanageable data quality and accuracy.



The lack of standardization and a single authoritative source of the truth makes complying with regulatory requirements almost impossible

## Root causes of poor data quality and accuracy



**Lack of industry standards** and data definitions for provider data and its exchange between payers and providers result in high complexity of data processing, making data quality vulnerable.



**Inaccurate data submission** by providers due to siloed data sources and a manually intensive process to extract and transform to custom formats, result in data integrity issues with source data.



**Lack of integrated workflow and real-time data integrations** between payers and

providers results in errors, data latency, manual reconciliation and multiple sources of truth.



**Lack of data strategy, governance and oversight** in both payer and provider data management systems result in poor handling of provider data, which is often contextual.



**Lack of data verification and enrichment** through primary sources and industry solutions, especially for delegated provider networks.





**Provider data is constantly changing**, making it a moving target without a centralized authoritative source of truth.



## The impact of poor-quality provider data is multi-faceted

Provider data is critical to ensuring the right level of service is provided to members and providers during their journey to find, manage and get paid for the care delivered. Key impacts of poor data quality on downstream business processes include:

 **Inaccurate data in the provider directory** impacts access to care and customer experience issues. High claims fall out rate can result in excessive cost of operation and provider dissatisfaction.

 **High claims fall out rate** can result in excessive cost of operation and provider dissatisfaction.



**High turnaround time** to onboard providers to the health plan network.



**Incorrect assignment of PCP** and member attribution to value-based care contracts.



**Care coordination and care outcome issues** due to inaccurate authorization processing.



**Regulatory non-compliance** resulting in sanctions and penalties.



**Poor user experience** due to manual data entry, processing, and reconciliations.



**55%**

Industry average for directory accuracy



**25 – 30%**

Average claims fallout rate



**30 – 90 Days**

Avg. onboarding time for providers



**~50%**

Provider data changes every 18 months

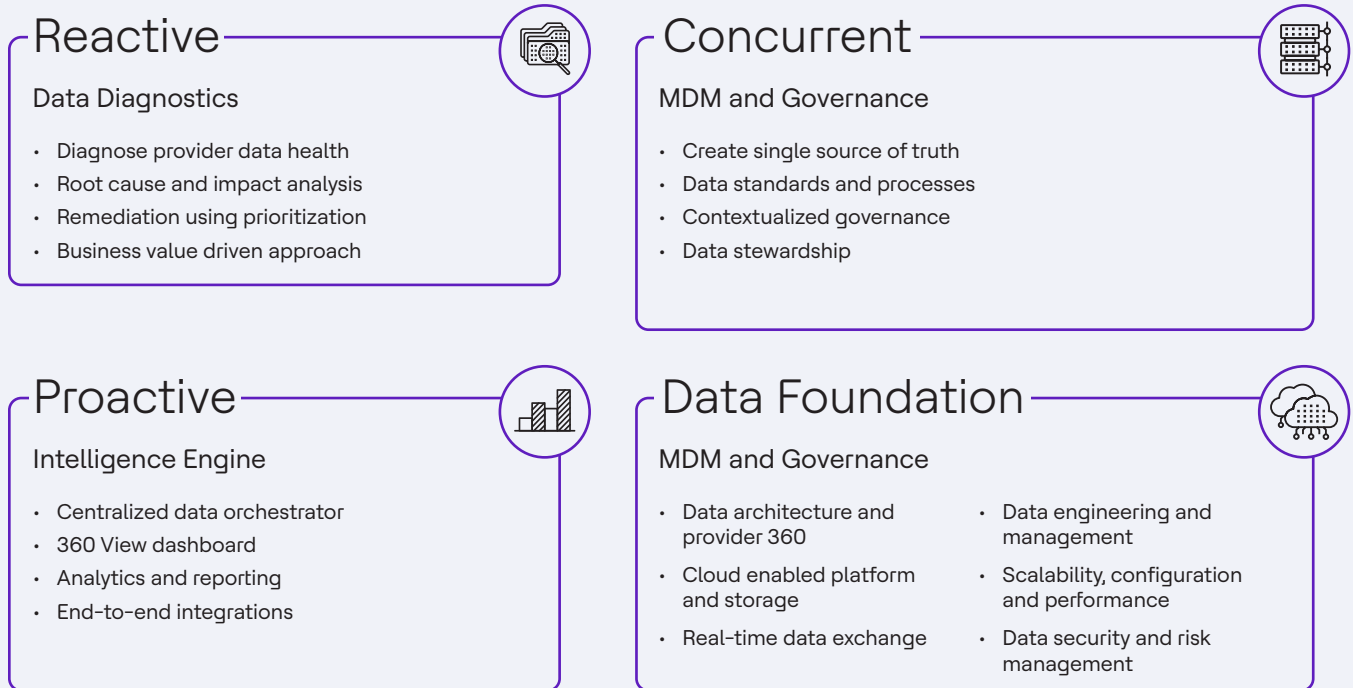


**\$8 to \$15**

PM cost to maintain data

# Solution Approach

HCLTech's four-pronged provider data strategy considers a futuristic data foundation with multiple layers of data management to ensure high-quality and accurate data is available for consumption. The plan addresses the need to remediate inferior quality data (reactive), objective time evaluation of data quality as it is being processed (concurrent) and predictive analysis that identifies the propensity of data to deteriorate in quality (proactive) as illustrated below:



The data strategy considers a futuristic data foundation with multiple layers of data management (reactive, concurrent and proactive) to ensure high quality and accurate data for consumption.

## Scalable and agile data foundation to enable provider 360 view

Many payers lack modularity, context and access to provider data that can be easily used for servicing or consumption. Fixing the data foundation means establishing the exemplary data architecture on the right foundations (cloud, API, data governance, configurability, performance, data security).

### The right data architecture

At the heart of a strong provider lifecycle management solution is the ability to build a Provider 360 architecture, which should enable drilled-down views and reports that are easy to produce and analyze and can initiate the next best actions for users and providers. Our solution creates a modular, multi-tier architecture, flexible schema evolution design and data aggregation at three levels (business entity, medical group and practitioner) so that various interrelationships can be explored easily in an interactive manner.

### The right foundation



**Cloud-based**—Migrate PLM applications and databases (including Provider SOR) to a cloud-based foundation to scale business operations. With initiatives like value-based care, telehealth and social determinants of health going mainstream, the need to unlock value from data is rapidly increasing, and legacy systems are unable to provide necessary business outcomes.



**API-first approach**– Integrate with source and surround systems in real-time through an API-first approach. With increasing ecosystem complexity due to the fast proliferation of diverse technology stacks, API lifecycle management is required to mitigate risks associated with data security, governance and cost escalation.



**Data engineering and management**– Reengineer provider data lifecycle to optimize cost and ease of maintenance with an adaptive data platform, reusable data ingestion framework, low code.

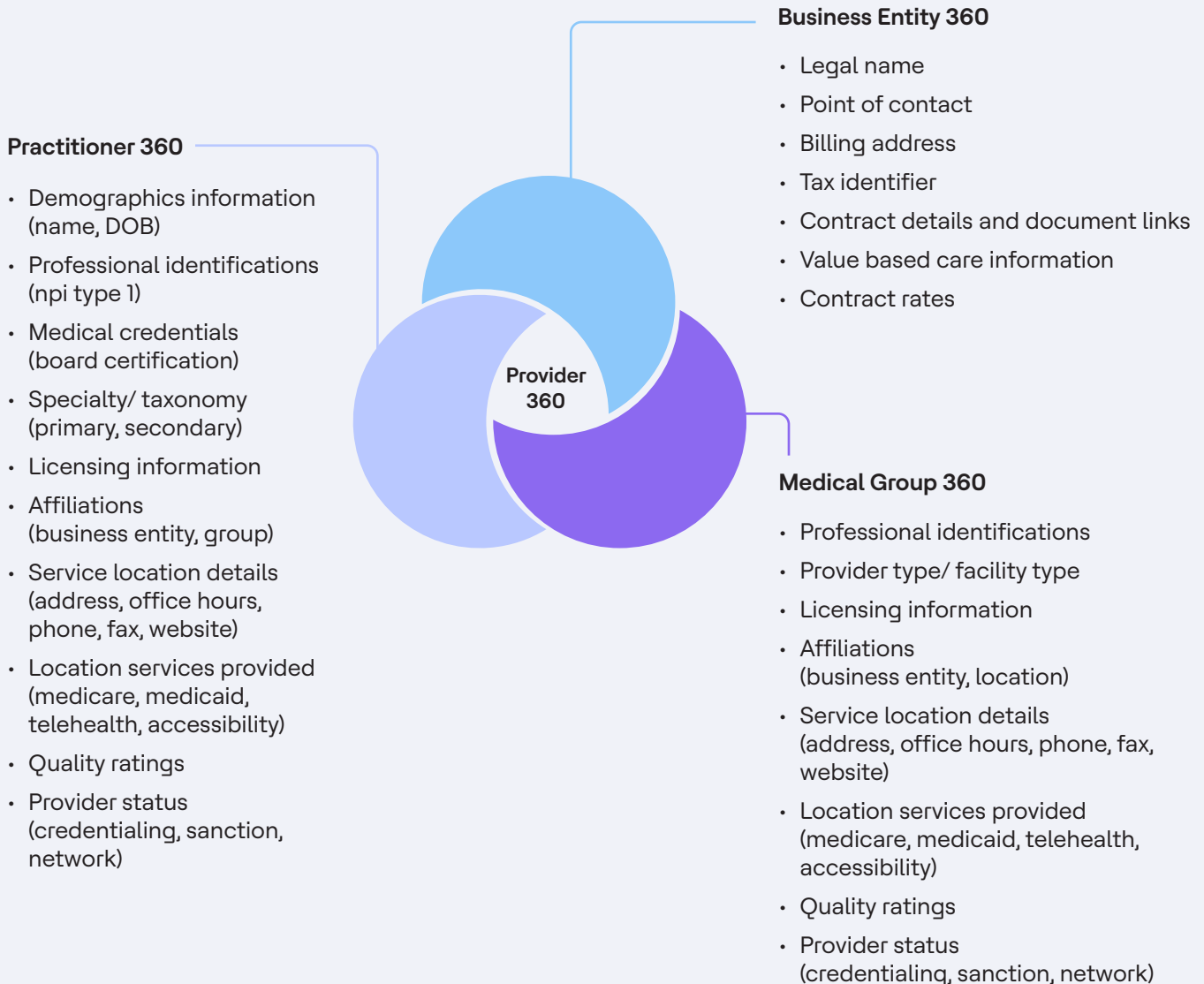
configuration driven data pipeline for data mapping/ transformation and new data source onboarding.



**Performance management**– Improve the availability of data for various consumption needs through continuous performance monitoring, making the system highly responsive to user actions at high load.



**Data security**– Identify sensitive PHI/ PII information, apply tokenization solution and role-based access controls.



## Data diagnostics to detect and cleanse bad data

The provider data universe comprises data that is generated or consumed during distinct phases of the provider lifecycle and impacts downstream processes very differently and with different intensities.

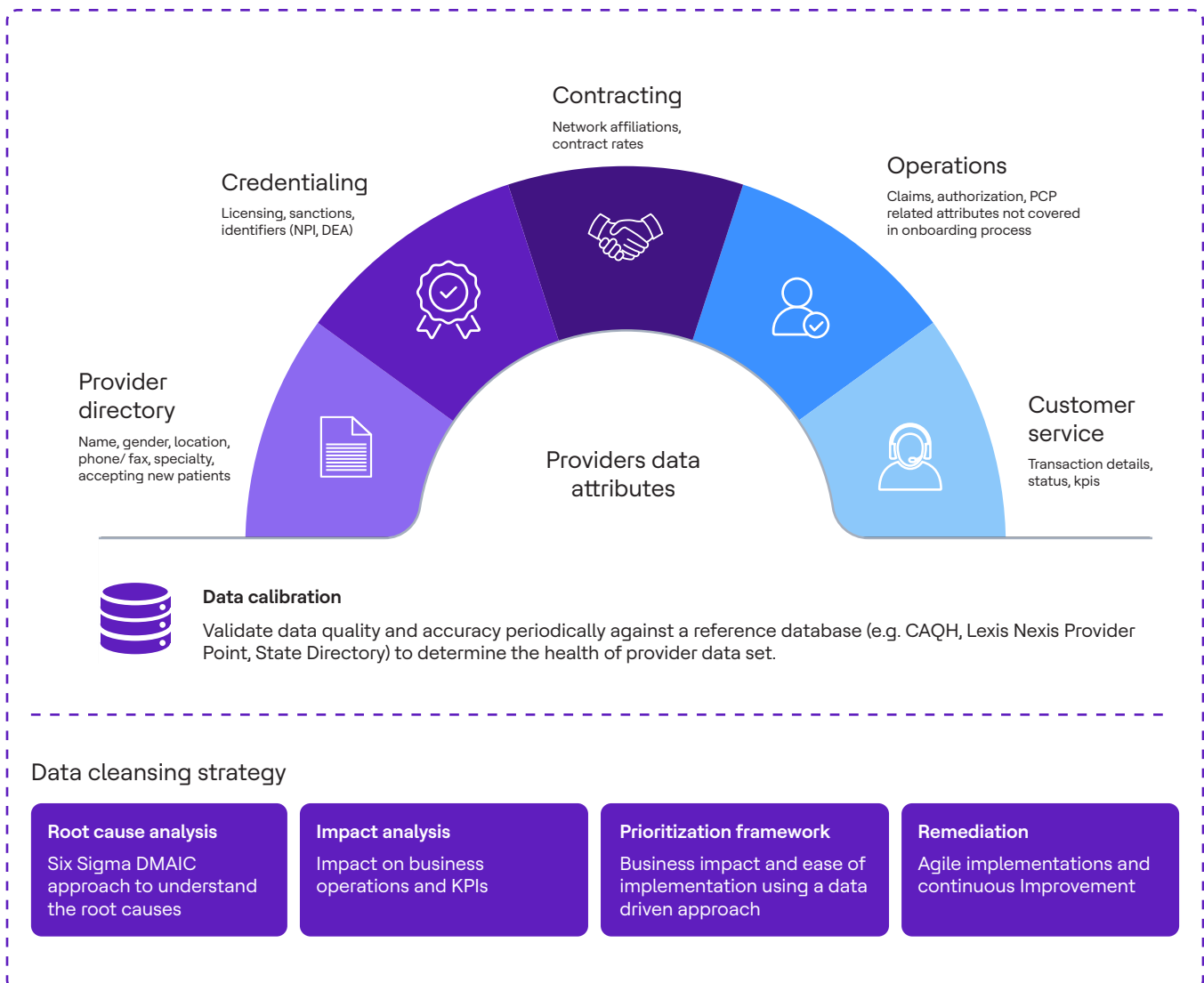
### Data segmentation

For an enterprise dealing with poor data quality, often the data cleansing efforts need to be prioritized based on measurable business impacts. With recent compliance focus, provider directory attributes have been the first to be prioritized. Similarly, other attributes need to be prioritized

based on qualification and quantification of impacted areas including provider onboarding, claims, authorization and customer service.

### Data calibration

The in-house provider data needs to be validated against an external reference source to determine the health of the provider data set. Many industry solutions exist that could help calibrate the in-house provider data to understand the data that needs to be retained, cleansed or enriched. The current vision of NDH will make it the source of truth for provider directory data against which payers will be able to cleanse their data and keep their systems concurrent.



## Data cleansing strategy

Adopt an ongoing strategy for data cleansing and institutionalize the governance processes using the following approaches:



**Root cause analysis**– Use of Six Sigma techniques like the DMAIC approach to understand the root causes of bad provider data. This should help initiate immediate fixes and permanent solutions and process changes that must be incorporated.



**Impact analysis**– Understand the impact of data cleansing efforts on business operations and KPIs, especially for long-term remediation, which could evolve into a project of its own.



**Prioritization framework**– Using a data-driven approach, quantify the business impact to build a high-priority and long-term roadmap of changes.



**Remediation**– Initiate agile-based implementations for incremental business value and continuous improvement of the process.

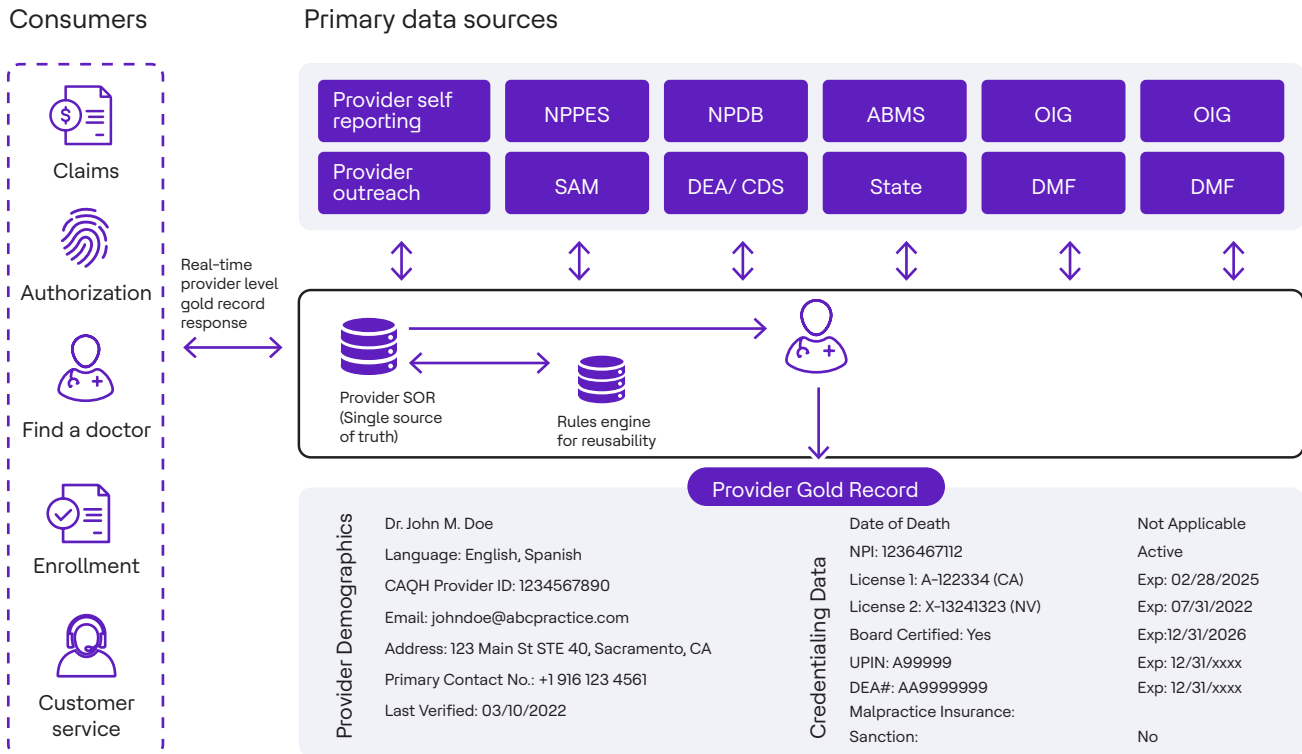
The data cleansing strategy will help the payer organization attain certain acceptable level of data quality but sustaining that level of quality requires additional checks and balances as provider data is very dynamic in nature as providers move, change jobs, add more qualifications or get sanctioned.



## Master data management and governance for sustainable data quality

The provider lifecycle datasets are generated and processed by multiple systems – external sources like provider or health systems as well as internal systems like provider network management. In addition, payers also use different data enrichment sources such as CAQH, NPPES, AMA, Lexis Nexis, State Files etc. Often, the data sets across these sources overlap, creating a need to prioritize the most trusted source in the given context. This requires building a governance framework that keeps track of data in real-time and certifies it as a 'gold record' as illustrated below:

Credential one provider once in a cycle, no matter how many requests originate from same or different PO for that provider



### Single source of truth

Establish hierarchy for data source and assign score/ weight to confidence level that it is still accurate.

If CL > Threshold = Gold Record

### Reuse gold record across PO requests

- Build a configurable rules engine that captures reusability rules across multiple categories like - state, PO, product, attribute specific.
- Prior to initiating primary source verification through external sources, verify if Gold copy for that provider already exists and reuse.

### Proactive expiration and sanction monitoring

- PSV request is triggered proactively to a specific source if dynamic attribute has not been verified for threshold no. of days or expiration is approaching.
- Monitor sanctions on a fixed interval and create alerts for further investigations/data sharing with POs.

### Sample illustrations:

# 1. If an attribute has a match in Gold record, and there are no conflicts, reuse the result from Gold record e.g, Date of Birth match

# 2. If an attribute is static (not expected to change) e.g. NPI, retrieve this information from inhouse Gold record and do not call NPPES every time a request is received for this provider either by different PO or for different network.

# 3. If an attribute can change but it is maintained say once a month, do not need to initiate verification if threshold days have not crossed. E.g. License information. Use the info from Gold record.



**Establish a source of truth** through a well-defined certification process that identifies a gold record at each provider entity level (business entity, practitioner, medical group).



**Identify a hierarchy for data sources** at every data element level and build a configurable rules engine that captures reusability rules across consumption use cases.



**Measure the level of accuracy** based on business logic when multiple sources are involved by assigning confidence level (CL) scores. Based on the CL score and ranges, the record could be qualified as Gold Record, Bad Record or Investigational Record.



**Reuse of gold records** to initiate data enrichment through external sources or business logic to avoid duplicative effort.



**Adopt a multi-domain master data management** approach to deliver standard, consistent high-quality data for consumption.



**Automated data lifecycle management** with the ability to initiate self-healing when data discrepancy is identified.



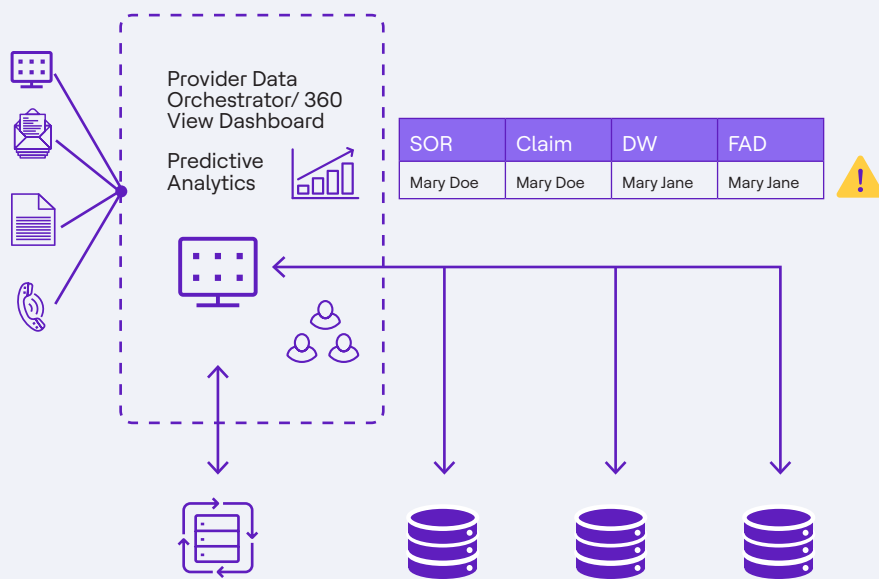
**Integrated workflow** to manage data discrepancies that cannot be managed through automated processes and make them visible on Provider 360 as alerts.



**Initiate data maintenance** at record level and bulk records via real-time data integrations with the Provider SOR.

Provider data quality is under scrutiny with multiple regulations adopted to improve care outcome and experience for members. This requires a fresh look at how data moves across multiple systems, gets processed and reported.

Data updates must reach directory within SLA of 48 hours<sup>1</sup>



### OUR POINT OF VIEW

Need for an intelligence engine that performs the following:

- **Centralized channel intake** and business rules management
- **Real time updates** to Provider SOR (granular and bulk updates)
- **Provider 360 view** with data comparison across systems
- **Analytics Engine** creates alerts for data discrepancies
- **Automated self cleansing** mechanism to repair bad data
- **Self service reports** – audit, operational, analysis, compliance, SLA reports

While the current industry focus is on provider directory data set, this design can be scaled to non-directory data attributes using a prioritization approach.

## An intelligence engine that drives proactive insights and self-cleansing

We envision a provider data management system that proactively evaluates the data quality, which is highly susceptible to changes, and initiates self-healing based on business rules and machine learning. The key components of HCLTech's intelligence engine will help payers perform following:



**Centralized channel intake** to a provider data orchestration system (Provider 360 view).



**Configurable business rules management** to dynamically add or update business rules.



**Real-time updates** to provider SOR, including granular and bulk updates.



**Provider 360 view** for longitudinal view of data and real-time comparison across enterprise systems.



**Analytics engine** that creates alerts for data discrepancies and engages business users proactively.

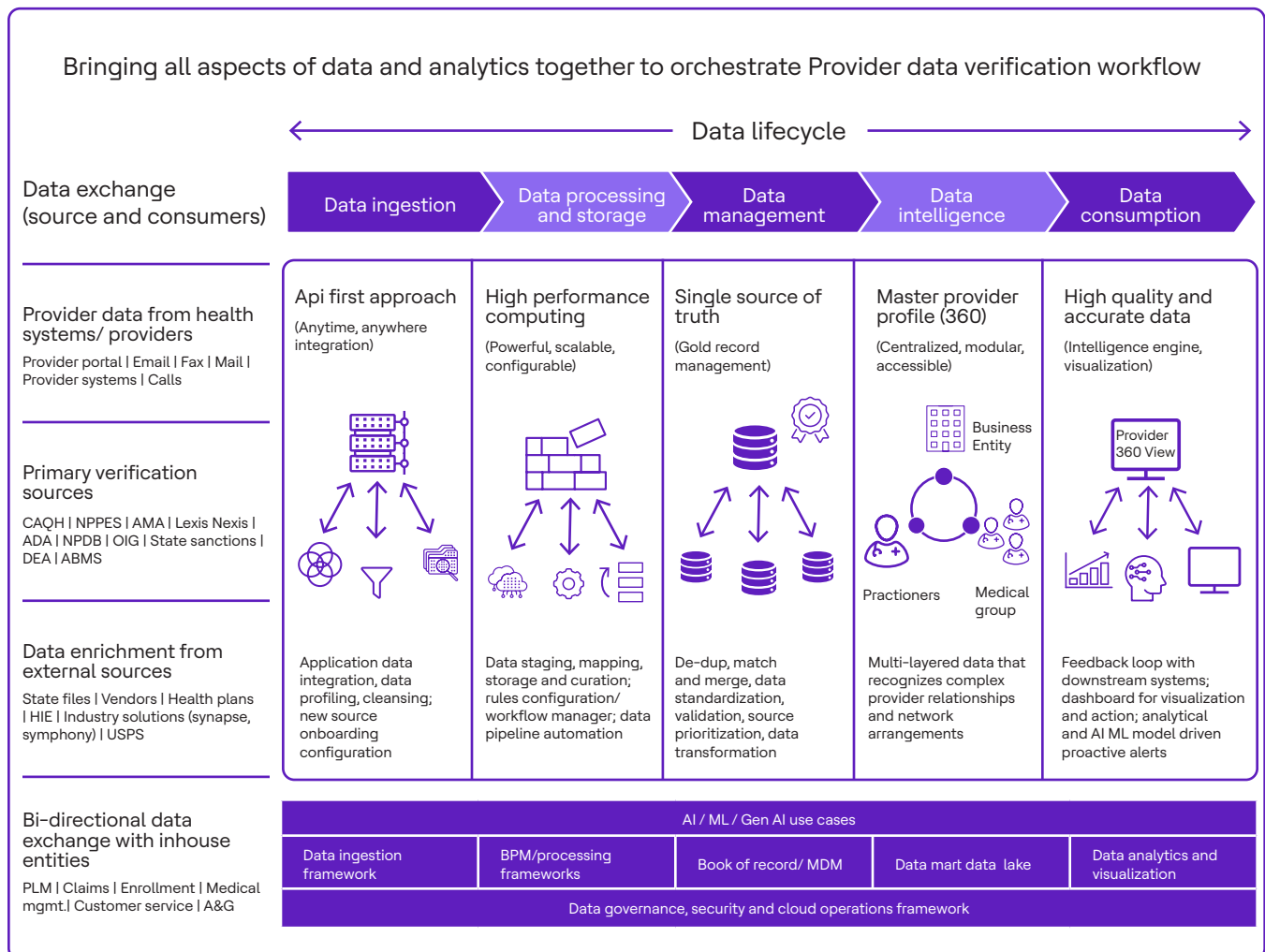


**Automated self-healing** mechanism to repair bad data and update algorithms for future corrections.



**Self-service reports** – Audit, operational, analysis, compliance and trend reports of initiated corrections.

Over a period, the engine will leverage machine learning algorithms to initiate self-healing based on similar corrections taken and further use that to identify vulnerable records.



## Bringing it all together to deliver a seamless provider data experience

HCLTech's holistic provider data strategy seeks to reorchestrate the provider data lifecycle from the source to consumption. This requires multiple technology solutions including data engineering, automation, workflow management, AI and analytics and powerful visualization to come together in an integrated manner.

The below provider data lifecycle framework illustrates how various components of our solution work in tandem to create an adaptive data platform that can deliver desired outcomes. The data lifecycle starts with ingestion of data from multitude of source systems (internal or external), its processing on a scalable foundation, data governance and making it available for consumption in proactive or on need basis.

AI and more specifically generative AI (GenAI), holds vast potential for application throughout the provider data lifecycle. Some illustrative examples of how GenAI can significantly enhance efficiency and reduce costs for health plans include:



**Data ingestion:** Convert non-standard, unstructured provider data in the form of emails, paper forms etc. into structured and standardized content with GenAI based data entry automation.



**Data processing:** Automate the data cleansing process for non-standard data to the required form and format and provide data mapping recommendations.



**Data management:** Identify inconsistencies, duplicates, and invalid data entry, detect trends and patterns in data discrepancies and apply to a broader set of affiliated data.



**Data intelligence:** Initiate self-heal or next best actions to help improve provider data quality and accuracy, identify providers at risk of attrition and build 360 views across application ecosystems.



**Data consumption:** Provide status on provider credentialing/onboarding, generate ad hoc conversational reports, develop insights from data based on user search and generate detailed provider profiles leveraging multiple sources.

The end goal is to create a resilient, scalable and flexible data foundation that can support collaboration with providers and drive customer experience by exchanging data in real time. If done right, the advent of modern technologies like GenAI will further drive productivity and quality improvements.



# Conclusion

The healthcare industry needs a comprehensive approach to solve challenges associated with provider data quality and integrity issues. The sole focus of government regulations on provider directory is akin to treating a symptom without fixing the root cause. While more needs to be done to establish the right standards and industry-wide governance to reduce the administrative burden on payers and providers, there is a strong need to build a cohesive data strategy for sustainable data quality and accuracy and investments into:



**Data foundations-**  
a scalable and agile  
data platform



**Data diagnostics-**  
ability to diagnose  
the health of data



**Data governance  
and MDM -**  
continuous  
monitoring and  
governance



**Intelligence  
engine-** layer  
of intelligence  
through AI/data  
science to derive  
proactive insights

HCLTech offers provider lifecycle management solutions that can be implemented with a modular approach that is well aligned to compliance requirements, recognizing the on-the-horizon initiatives to address the broader industry needs. We are continuously infusing GenAI into our platform solutions and have a wide variety of use cases that can be applied to custom client environments.

# References

- CMS (Oct 05, 2022), Public Input on Establishing First, National Directory of Health Care Providers and Services from <https://www.cms.gov/newsroom/press-releases/cms-asks-public-input-establishing-first-national-directory-health-care-providers-and-services>
- Sarai Rodriguez (Dec 07, 2022), Article: AHA Asks CMS to Halt National Healthcare Provider Directory Operation from <https://ehrintelligence.com/news/aha-asks-cms-to-halt-national-healthcare-provider-directory-operation>
- HL7 International - Patient Administration Work Group (2022), HL7 National Directory of Healthcare Providers and Services (NDH) Implementation Guide 0.1.0 – draft from <https://build.fhir.org/ig/HL7/fhir-us-nadh/>
- CAQH and AMA (2022, January). Improving Health Plan Provider Directories and the Need for Health Plan-Practice Alignment, Automation and Streamlined Workflows from [https://www.caqh.org/sites/default/files/other/CAQH-AMA\\_Improving%20Health%20Plan%20Provider%20Directories%20Whitepaper.pdf](https://www.caqh.org/sites/default/files/other/CAQH-AMA_Improving%20Health%20Plan%20Provider%20Directories%20Whitepaper.pdf)
- CMS (2018, January 31). Online Provider Directory Review Report by CMS from [https://www.cms.gov/Medicare/Health-Plans/ManagedCareMarketing/Downloads/Provider\\_Directory\\_Review\\_Industry\\_Report\\_Round\\_2\\_Updated\\_1-31-18.pdf](https://www.cms.gov/Medicare/Health-Plans/ManagedCareMarketing/Downloads/Provider_Directory_Review_Industry_Report_Round_2_Updated_1-31-18.pdf)
- CMS (April 6, 2015), Announcement of Calendar Year 2016 Medicare Advantage Capitation Rates and Medicare Advantage and Part D Payment Policies and Final Call Letter.
- CMS (2016), Advance Notice of Methodological Changes for Calendar Year (CY) 2016 for Medicare Advantage (MA) Capitation Rates, Part C and Part D Payment Policies and 2016 Call Letter (cms.gov)
- Lexis Nexis Provider Data Masterfile Solution - High-quality provider data is foundational to provider communications, informed patients and successful decision making from <https://risk.lexisnexis.com/products/provider-data-masterfile>
- CAQH Proview for Groups Solution from <https://www.caqh.org/solutions/caqh-proview-groups>
- Sokol, E (2020, September 28). Value-based care stunted by misaligned payer-provider relationship from <https://healthpayerintelligence.com/news/value-based-care-stunted-by-misaligned-payer-provider-relationship>
- OCXCognition (2020). NPS Benchmark Analysis for the Health Insurance Industry from [https://ocxcognition.com/wp-content/uploads/2020/07/ocxcognition\\_NPS\\_Benchmark\\_Analysis\\_for\\_the\\_Health\\_Insurance\\_Industry.pdf](https://ocxcognition.com/wp-content/uploads/2020/07/ocxcognition_NPS_Benchmark_Analysis_for_the_Health_Insurance_Industry.pdf)
- CMS (2022, March). Medicare Provider Data Catalog – Doctors and Physicians from <https://data.cms.gov/provider-data/topics/doctors-clinicians>
- CAQH and Manatt Health (2016) White Paper: Defining the Provider Data Dilemma from <https://www.caqh.org/explorations/white-paper-defining-provider-data-dilemma>
- Synaptic Health Alliance White paper (2019) on Improving Provider data accuracy by synaptic health alliance from <https://www.synaptichealthalliance.com/>
- Symphony California's State Provider Directory Solution from <https://www.iha.org/provider-directory-management/symphony-provider-directory/>

# HCLTech | Supercharging Progress™

HCLTech is a global technology company, home to more than 219,000 people across 60 countries, delivering industry-leading capabilities centered around digital, engineering, cloud and AI, powered by a broad portfolio of technology services and products. We work with clients across all major verticals, providing industry solutions for Financial Services, Manufacturing, Life Sciences and Healthcare, Technology and Services, Telecom and Media, Retail and CPG, and Public Services. Consolidated revenues as of 12 months ending June 2024 totaled \$13.4 billion. To learn how we can supercharge progress for you, visit [hcltech.com](https://hcltech.com).

[hcltech.com](https://hcltech.com)

