The Case for FHIR-based Quality Measurement and Reporting

# Overview

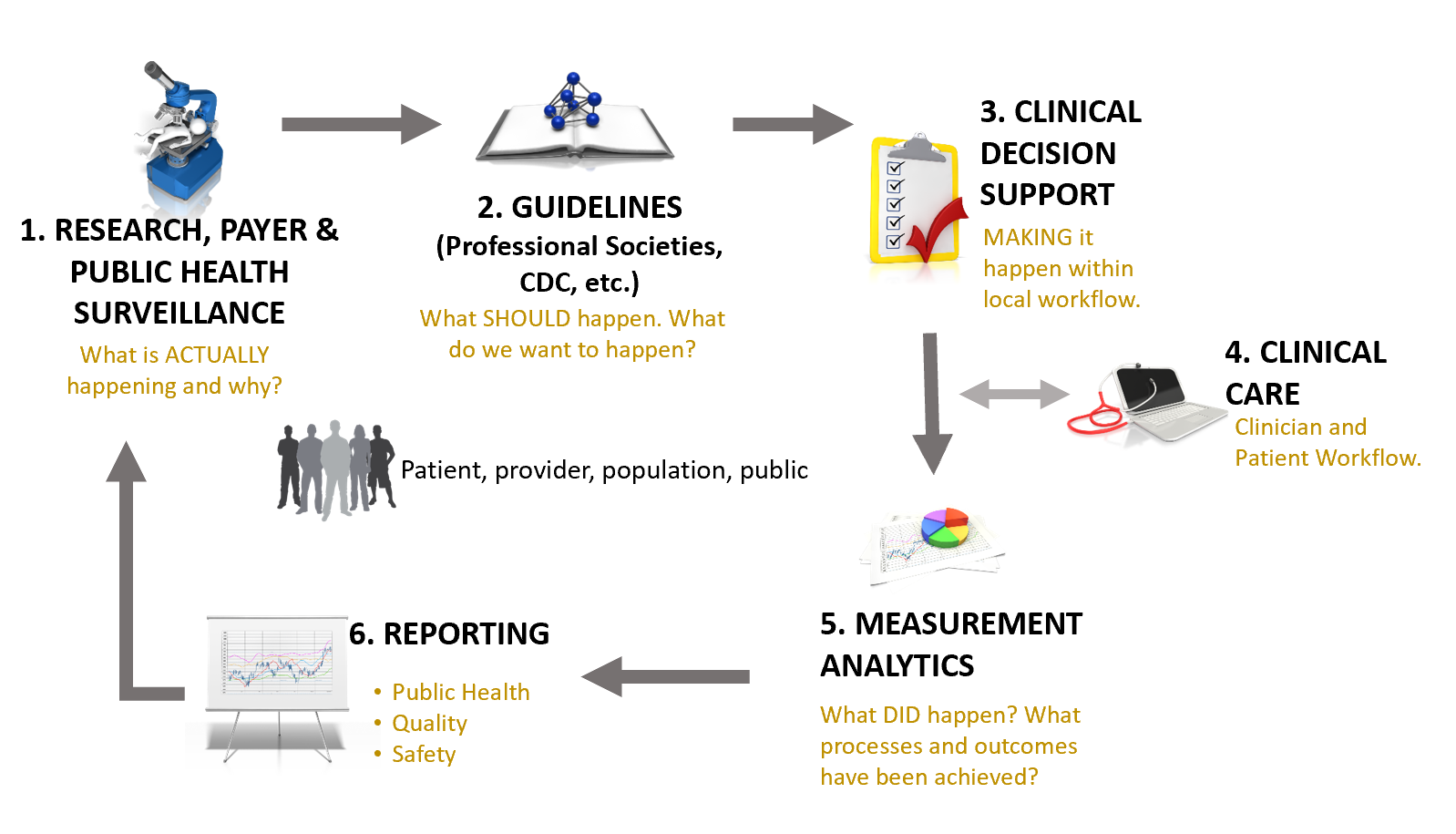
There has been an increased emphasis on care quality measurement over the last decade. The United States pays a high amount for healthcare yet falls behind other nations based on quality, morbidity and mortality. The federal government along with medical societies, quality measurement organizations and payers are focusing more on value for quality-based care. To this end, clinical measures have been created which can be used as part of a “virtuous cycle” of continuous quality improvement.

This white paper makes the case that using Fast Healthcare Interoperability Resources (FHIR) for quality measurement specification, distribution, evaluation, and reporting provides many benefits that enable positive change in the clinical quality improvement ecosystem, including reducing reporting burden, increasing the accuracy and fidelity of reporting results, shortening the reporting cycle, and accelerating the overall clinical quality improvement lifecycle.

# Background

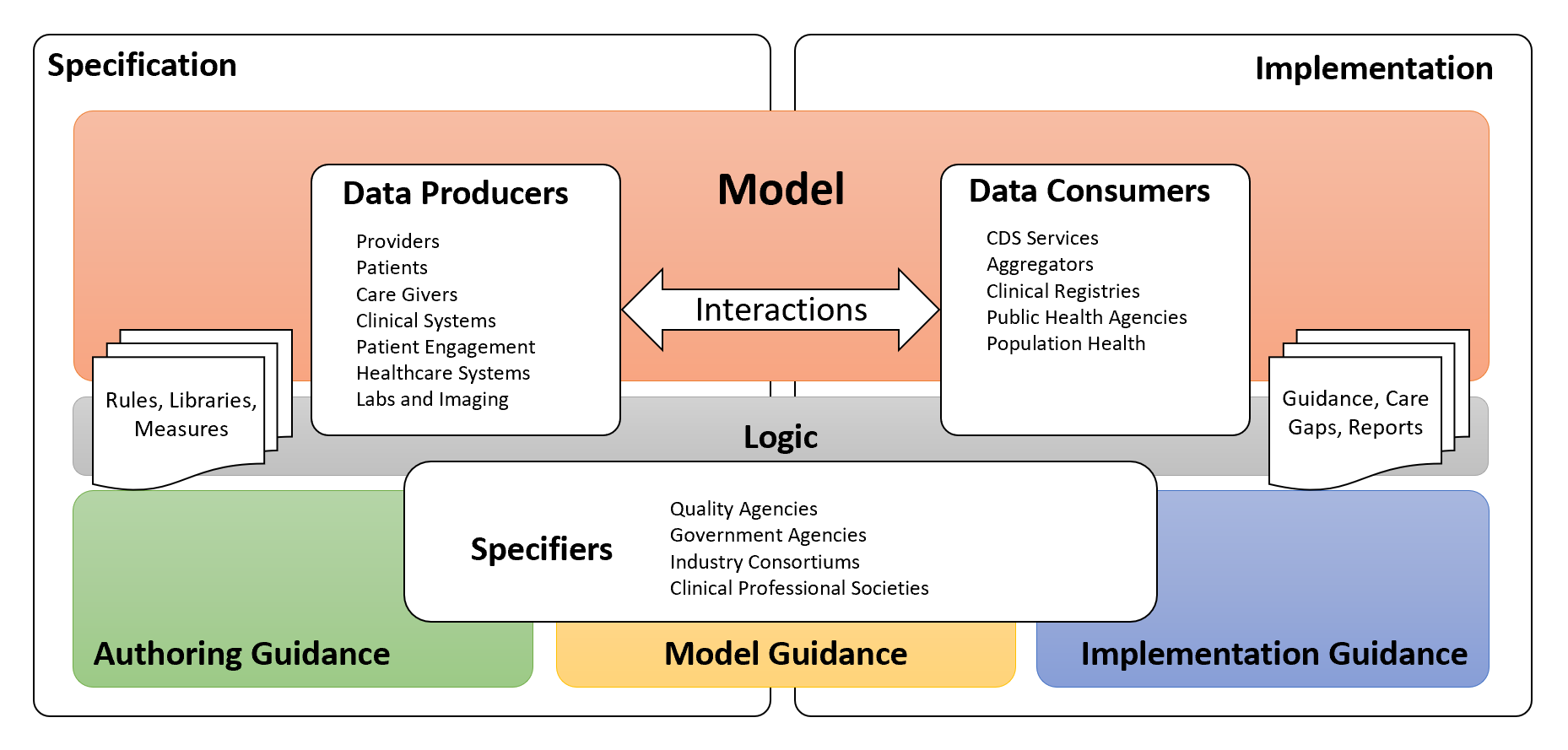
Since their inception in 2009, the Quality Data Model (QDM) and Health Quality Measure Format (HQMF) have provided a conceptual basis for the specification of electronic Clinical Quality Measures (eCQMs). Beginning in 2014, the Clinical Quality Framework Initiative was launched to identify, develop, and harmonize standards that promote integration and reuse between the broad domains of Clinical Decision Support and Clinical Quality Measurement. Specifically, to enable knowledge interoperability, from discovery to delivery and back, and at scale, the initiative focused on defining a platform and model independent mechanism for sharing the logic required to support CDS and CQM use cases. The resulting logic specification, Clinical Quality Language (CQL), allows logic to be expressed and shared using a specified data model bound to standard terminologies. CMS is currently using CQL to distribute eCQM specifications, with QDM as the data model, bound to standard terminologies including ICD, CPT, SNOMED, RxNorm, and others. One of the primary motivations for identifying separate specifications for the logic and data model was to allow the logic and data model specifications to evolve independently, and enable flexibility in selecting a data model. Furthermore, by focusing on evolving and separating the logic specification first, implementations could focus on updating their engines or translation paths while holding the current data model (QDM) relatively constant under that change. This architectural approach further enables more flexibility in adopting data model changes, up to and including entirely different data models. As the standards landscape continues to evolve, it has become clear that FHIR has reached critical mass, and that there are clear opportunities to reduce transformation and implementation burden and complexity, while simultaneously enabling richer knowledge interoperability use cases, across the full clinical quality improvement lifecycle, and specifically for quality measurement.

To set this in the larger context, this diagram represents the overall Quality Improvement Ecosystem:

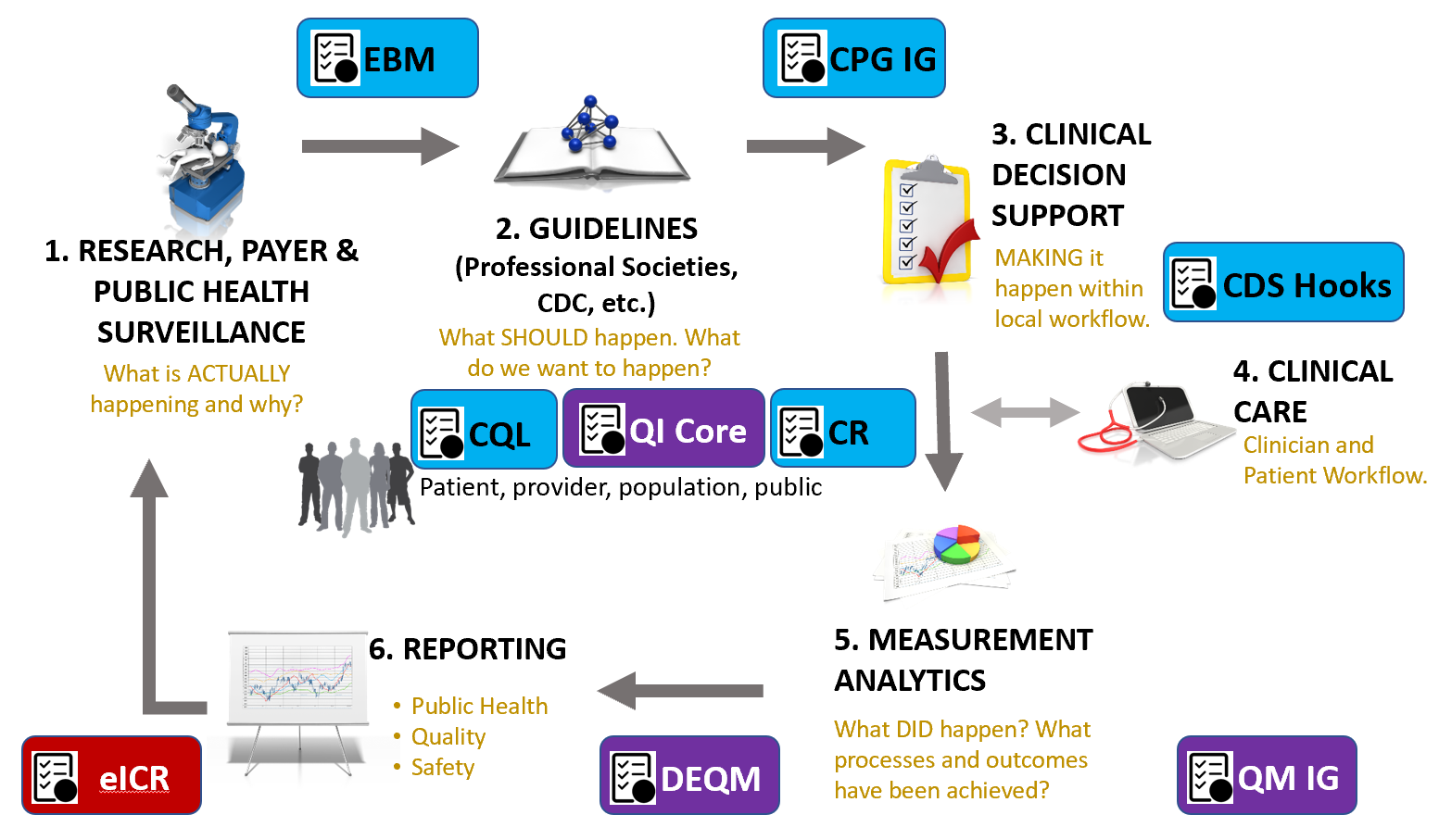


As shown in step 1, this ecosystem begins with information and knowledge discovery, preferably evidence-based from formalized research, public health surveillance, and data science and other analyses performed by third parties such as payers and organizations informing policy. Such information indicates the existing state of knowledge about a given clinical topic. In step 2, stakeholders, such as professional societies, public health and governmental bodies, and industry consortiums and conveners have various methods for disseminating this information to assure awareness among consumers, healthcare practitioners, and healthcare organizations about what is known and suggested methods for managing the clinical topic. Ideally, suggested or recommended practices are captured and documented in guidelines based on collaboration among recognized clinical domain experts and clinicians, clinical research and guideline development experts, patient and consumer advocates, and more recently informaticians and terminologists with both knowledge translation and clinical workflow and implementation expertise. In step 3, these clinical guidelines are translated into clinical decision support (CDS) artifacts to incorporate valuable clinical recommendations and actions directly within clinical workflow and to provide cognitive support for effective clinical decision-making. To adequately impact clinical care for clinicians and patients requires local practice-level implementation activities as shown in Step 4. Ideally, in Step 5, the clinical guidelines and resulting CDS include methods or metrics for evaluating what successful implementation means, i.e., whether the clinical care ultimately delivered recommended care process activities that addressed the intent of the guideline and if it achieved the desired end or intermediate outcomes with consideration for expected (codified in logic) or expressly communicated (documented) exceptions. In Step 6, to close the loop and enable continuous improvement, the results of such measurement analytics, supporting raw data and enriched information must be reported for detailed and aggregate review. This aggregate and detailed data serves the purpose of evaluating clinical performance, patient and care team details for continuous improvement, and key outcomes for healthcare organizations and other relevant parties (e.g. public health and payers). Detailed and related raw clinical data, key enrichments (e.g. situationally valuable inferences, new or ensuing clinical documentation), and information (or metadata) on clinical activities and local factors (e.g. CDS response rate, site of service setting and location).

Shown another way, the following diagram depicts the various stakeholders and interactions between them that make up this ecosystem:



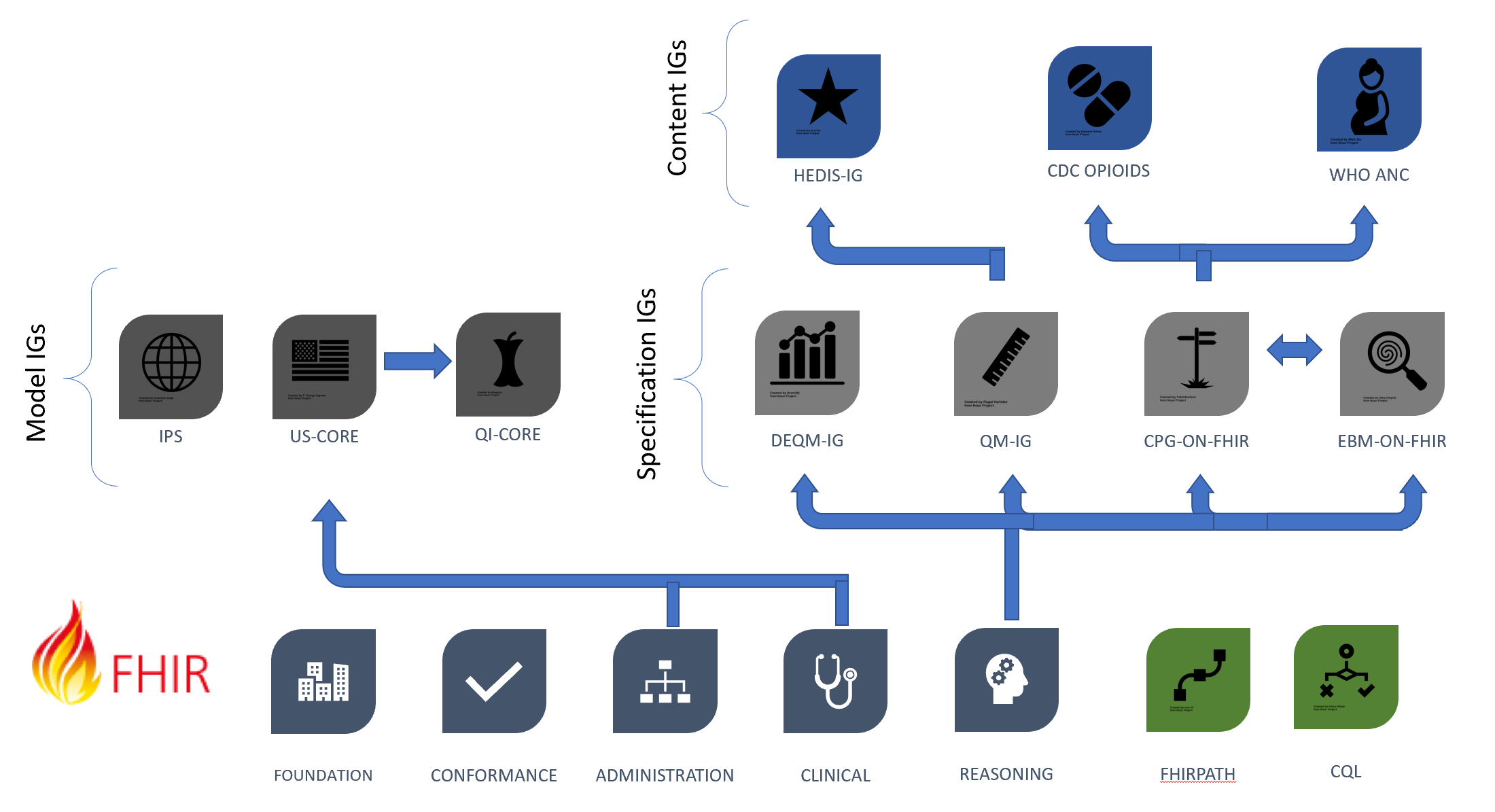
Standards and specifications enable the precise description of and guidance for the various use cases found throughout this ecosystem. In particular, many of the projects in the Clinical Quality Information and Clinical Decision Support work groups are focused on supporting quality improvement use cases:



Specifically:

|  |  |
| --- | --- |
| [Evidence-Based Medicine-on-FHIR](https://confluence.hl7.org/display/CDS/EBMonFHIR) | This project specifies resources and patterns for the exchange of data involved in evidence-based medicine including study results, quality of evidence and strength of recommendation (e.g. GRADES) and relevant context (e.g. PICOT), environmental surveys, and systematic reviews. |
| [Clinical Practice Guidelines on FHIR](https://confluence.hl7.org/display/CDS/CPGonFHIR) | This project supports the use of FHIR resources to build shareable and computable representations of the content of clinical care guidelines and recommendations. |
| [Clinical Quality Language](https://confluence.hl7.org/display/CDS/Clinical+Quality+Language) | This project defines a representation for the expression of clinical knowledge or logic that can be used across the quality improvement ecosystem. |
| [Clinical Reasoning](http://hl7.org/fhir/clinicalreasoning-module.html) | The Clinical Reasoning module provides resources and operations to enable the representation, distribution, and evaluation of clinical knowledge artifacts such as clinical decision support rules, quality measures, public health indicators, order sets, and clinical protocols. |
| [Quality Improvement Core](http://hl7.org/fhir/us/qicore/) | The QI-Core Implementation Guide defines a set of FHIR profiles with extensions and bindings needed to create interoperable, quality-focused applications. The profiles in this implementation guide derive from and extend the [US Core](http://hl7.org/fhir/us/core) profiles to provide a common foundation for building, sharing, and evaluating knowledge artifacts across quality improvement efforts in the US Realm. |
| [CDS Hooks](https://cds-hooks.hl7.org/) | The CDS Hooks specification describes the RESTful APIs and interactions to integrate Clinical Decision Support (CDS) between CDS Clients (typically Electronic Health Record Systems (EHRs) or other health information systems) and CDS Services. |
| [Quality Measure IG](http://hl7.org/fhir/us/cqfmeasures/) | The QM IG describes an approach to representing electronic Clinical Quality Measures (eCQMs) using the FHIR Clinical Reasoning Module and Clinical Quality Language (CQL) in the US Realm. |
| [Data Exchange for Quality Measures](http://hl7.org/fhir/us/davinci-deqm/) | The DEQM IG supports quality reporting and data exchange for value-based care in the US Realm. |

The following diagram illustrates the connections and relationships between the FHIR-based specifications and guidance discussed above:



The bottom row depicts the foundational standards of FHIR, FHIRPath, and CQL. On the left side of the middle row are *model implementation guides*, or implementation guides that focus on enabling semantic interoperability for clinical data for a broad range of use cases. On the right side of the middle row are the *specification implementation guides*, or implementation guides that provide conformance expectations and guidance for building shareable and computable content using the FHIR Clinical Reasoning module, focusing on different use cases within the quality improvement ecosystem. And finally, in the top row are *content implementation guides*, or implementation guides published by domain authorities (such as professional societies, government health and regulatory agencies, healthcare institutions, and others), that provide computable content, conforming to a specification IG, for specific use cases such as a quality measurement program, or a clinical guideline.

Within this larger context, this white paper identifies the benefits, and highlights the challenges, associated with the use of the FHIR-based quality measurement and reporting specifications defined as part of this overall quality framework. More specifically, this paper makes the case that although there are benefits to identifying, building, and maintaining a quality measurement-focused model like QDM, the benefits of using the FHIR framework as the basis for Quality Measurement use cases outweigh the collective advantages derived from a use-case-specific model.

# Summary

The following table summarizes the main points made throughout the paper:

|  |  |
| --- | --- |
| Expressivity | FHIR supports a broad range of use cases, including not only clinical data, but claims, labs, research, public health, and other applications from across the healthcare domain |
| Alignment | FHIR is developed with a focus on implementation, providing alignment with existing clinical information systems and clinical workflows, greatly reducing the burden of semantic transformation |
| Fitness | A common data model for logic expression and data transport/ exchange further reduces the opportunities for mis-aligning semantics between distinct models resulting in unintended inference/ interpretation. FHIR can serve as data model for both logic expression and data representation/exchange |
| Liquidity | FHIR is an API-based approach, enabling applications as well as reporting capabilities and knowledge assets to be portable and fungible |
| Community | FHIR is a diverse and dedicated community, bringing expertise and experience to bear not only on how the specification is built, but how it is used – particularly around clinical use |
| Extensibility | FHIR has a well-defined and flexible mechanism for supporting coordinated exchange of use-case specific information, without undermining core interoperability |
| Conformance | FHIR has a rich conformance framework for describing and validating exchanges and ensuring interoperability |
| Tooling | FHIR has a well developed and well supported set of open source and vendor tooling for authoring, modeling, developing, publishing, and implementing |
| Agility | FHIR has a rich set of publication tooling to support development and implementation of use cases through specifications, implementation guides, and supplements |
| Reusability | FHIR provides a foundation for sharing content both within and across use cases, increasing opportunities for reusable content across guidelines, decision support rules, quality measures, case reporting, and workflow applications |
| Implementability | FHIR directly supports expression of quality measurement use cases with the Measure and MeasureReport resources |

## Expressivity

Across the Clinical Quality Ecosystem, there is a need for fully and accurately expressing explicit meaning for data elements within logic expression as well as mechanisms for data exchange including transport, messaging, and reporting. Expressivity addresses the scope- breadth and depth- of information that can be represented in a data model. Variety (diversity, range), extent (specificity, fidelity), and quantity of concepts expressed are key criteria on which to assess the expressivity of a data model. Expressivity, Alignment, and Semantic Fitness are related concepts. Expressivity addresses whether explicit, unambiguous meaning can be sufficiently represented in a model. Alignment addresses closeness of data semantics across use cases, most importantly existing systems of record. Fitness addresses how closely data semantics convey or express real-world concepts.

As a clinical conceptual model, Quality Data Model has primarily (and initially exclusively) focused on clinical content that would be available within a patient’s medical record, and required for a specific set of quality measures. However, claims data is an important aspect of many types of healthcare quality measures, and more recent versions of Quality Data Model have begun to include basic coverage information. As a general-purpose framework for healthcare data exchange, FHIR supports a rich and expanding set of use cases, including payment, coverage, claims, and plan eligibility and enrollment. Quality measures that use FHIR can begin using these resources to more easily integrate clinical and claims data in the same measures to address critical use cases across clinical and financial spectrum as well as leverage key information from a broader set of data sources (clinical, claims, registry, etc.), often even in combination to more comprehensively address the information need (e.g. did a patient receive a critical service such as a colonoscopy, mammogram, or HbA1c lab).

As one example, the FHIR resources are richly interconnected, supporting not only the description of clinical and other patient-related data, but also the relationships between those data elements such as the performers of an Encounter or Procedure, the target of a Communication, or the beneficiary of Coverage. These relationships are a critical aspect of establishing relevant criteria in a quality measure, but have traditionally been expressed using timing relationships between the data elements, rather than direct references expressed in the information model.

In addition, quality measurement use cases involving both claims and clinical data (sometimes referred to as Hybrid Measures) are an important use case that can begin to be met with the administrative and payment-related resources in FHIR, specifically the Claim and Coverage resources to support describing insurance plan participation.

And finally, this richness supports the ability to formally define provider attribution criteria, whereas traditional approaches have relegated that to narrative descriptions in program guidance. This lack of formal representation has been a source of implementer burden and confusion, and recent efforts have begun to address this more formally; with FHIR, the data elements required to more fully express these criteria are available in currently published versions of the specification. Attribution is but one use case that clearly requires a combined perspective on clinical and financial/administrative data.

## Alignment

FHIR was primarily designed as a healthcare interoperability standard with a heavy focus on implementability. The FHIR development process has always included real-world implementation as part of developing, validating, and publishing the standard[[1]](#footnote-1). As a result, the FHIR data model is closely aligned with what existing clinical information systems support, especially for the more mature FHIR resources. This alignment reduces transformation burden, implementation and development effort, time, and cost, as well as the potential for error through semantic mis-alignment. The less transformation logic has to be written, the more fidelity the data exchanged will have, and the higher quality data available for use in all aspects of healthcare, including quality measurement.

In considering whether to continue development of a use-case specific conceptual model for quality measurement, this alignment question is key to reducing implementer burden. Enabling data exchange between systems requires a common model. The more different systems involved in the exchange, the more effort is required to transform data to and from that common model. In the case of FHIR, implementer systems are already performing this transformation from their internal data models to FHIR. Introducing another conceptual model on top of that introduces another layer of mapping and transformation that must be developed, tooled, authored, implemented, and subsequently maintained over time. And the more layers of mapping and transformation involved, the greater the chance for errors, semantic misalignment, and loss of fidelity.

## Fitness (Representational Bias)

Semantic fitting deals with the alignment of the semantics or meaning that can be conveyed, represented, or expressed in the data model. Overfitting, underfitting, and mis-fitting can all create unnecessary complexity and worse, inadvertent or even unexpected variation in the intended versus interpreted meaning of information resulting from the inferences of clinical business logic. Overfitting occurs when there is more extensive expressivity in the data model than is required to completely, accurately, and sufficiently represent domain entities and concepts within the information carrier (data model) resulting in ambiguity, variability, and misinterpretation. Underfitting occurs when a data model does not or cannot represent ample concepts to carry the information required to sufficiently describe the domain of interest. Mis-fitting occurs when there is simply misalignment between the source and target of representations of information across use cases and/or from domain entity/concept to the expression of information.

FHIR, and QI-Core, are designed to express the data semantics for the use cases of clinical care, delivery, and supporting financial processes as well as the full lifecycle of Clinical Quality Improvement respectively. This, in addition to the Extensibility and Community beneficial points listed below, holds much more opportunity and likelihood for much better “fitness” for semantic representation and ultimately valuable and actionable application of Quality Measurement and related CDS interventions and other CQI use cases such as Case Reporting and Knowledge Discovery.

## Liquidity

This same semantic alignment and appropriate fitting through FHIR also facilitates the re-use of knowledge assets developed for quality measurement across other high-value use cases such as integration with clinical workflow via CDS Hooks, or data enrichment and insight delivery via SMART-on-FHIR applications, surveillance and reporting to professional societies, payers, and research or government entities, or even for direct use within systems that support write access in their FHIR services.

For example, decision logic content that is developed from the same guidelines used to inform the quality measurement specifications can be operationalized at the point-of-care using applications that can deliver that content, driving performance improvement in support of quality measurement. Inclusion and exclusion criteria for measure populations often overlap with “Condition” criteria in common CDS Event-Condition-Action Rules, Eligibility criteria for clinical Pathways (derived from CPGs), and cohort definitions (“triggers”) and supporting data elements for eInitialCaseReports (eICRs). The opportunity for shared logic and data elements (raw and inferred) across these use cases greatly reduces unwarranted friction within the CQI lifecycle and likewise makes possible much more expedient best practice (knowledge) discovery-to-delivery translational and implementation endeavors.

In addition, the existence of API-based access to clinical systems has already been a driver for clinical quality and analytics use cases beyond quality reporting and measurement, and the types of exchanges enabled (such as incremental data submission throughout a given measurement period) are already driving standardization around care management processes such as Gaps in Care identification, closure, and potentially even prevention.

## Community

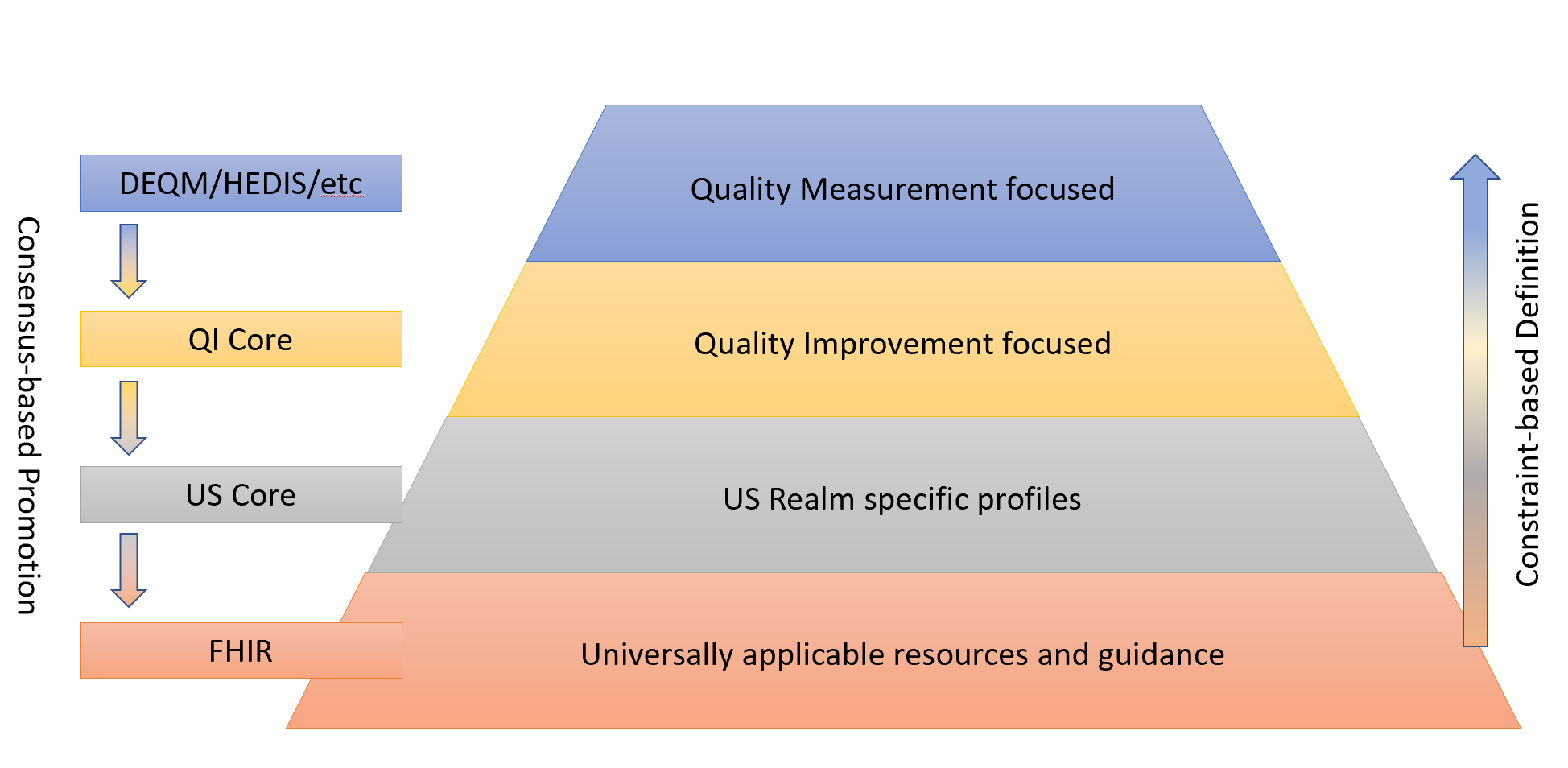
The FHIR specification benefits from a broad and ever-expanding community of stakeholders from across the healthcare industry[[2]](#footnote-2). As a focused data model, QDM currently has a representative set of stakeholders from within the quality measurement community, but as the use cases for quality measurement expand, so must that representation across the clinical care and healthcare delivery domain. By using FHIR, quality measurement joins the already vibrant community contributing to the shape and supported use cases of the FHIR resources and implementation guidance. This brings clinical practice and quality measurement in closer alignment in terms of data, information, and knowledge (business logic) assets as well as process, governance, and focus/prioritization.

As a specific example of the benefits of engaging the FHIR community, a recent measurement use case involves representation of the nutrition orders for a patient. However, the NutritionOrder resource within FHIR has a low maturity level, so measure developers were able to reach out to the stewards of the resource through the Orders and Observations HL7 Work Group to bring their use case to a forum for discussion. Specifically, the measurement use case involved ordering of a specific nutrient, exclusive of other nutrients, rather than an order for a diet, which is what the NutritionOrder resource supported. This discussion revealed that the NutritionOrder resource is being replaced by the NutritionIntake resource in future versions of FHIR, and that until future version is available, it makes more sense to model the required data element using a ServiceRequest resource in the current version of FHIR, and adapting that usage to the more specific resource once it becomes available. In addition, the specific use case requirements brought by the measure developers are now part of the conversation as the NutritionOrder resource is being evolved.

## Extensibility

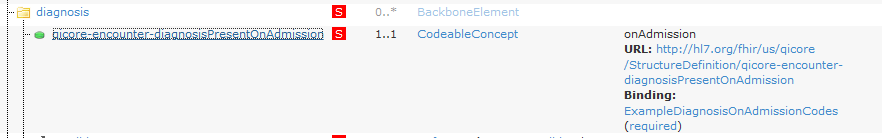
No matter how detailed the information model provided by the FHIR framework and associated implementation guides, various pressures including evolving use cases, regulatory and implementation environments, as well as advances in medical knowledge and the practice of care will require that the information model evolve over time to support additional content and concepts (semantics). Particularly, the ability to extend and constraint via profiling such as in QI-Core (or US-Core) enables adaptation of the data model to improve “Fitting” to the use cases and domain.

The FHIR framework provides a well-established and flexible mechanism for defining additional data elements, as well as a process for achieving consensus of those new data elements for inclusion at the appropriate level of exchange. The diagram below illustrates this landscape:



As new use cases are proposed, FHIR as a platform specification supports the ability to describe extended information that is not defined by the base standard. Using profiles and implementation guidance, that extended information is still specified in a formal way that enables conformance checking. Moving up the diagram involves constraint-based definition to capture the use-case specific requirements, and moving down the diagram involves submitting feedback to base implementation guides and specifications, and achieving consensus where appropriate for content that is more broadly applicable.

As an example of this process, some quality measurement use cases rely on identifying whether a particular diagnosis was present on admission. This information is not currently represented within the Encounter or Condition resources in FHIR. The QICoreEncounter profile defines an extension to represent this information:

Recognizing that this information is more broadly applicable than just quality measurement, this extension is being submitted as feedback to the USCore profiles, and ultimately the base FHIR specification itself to support capturing that data as a first-class data element.

## Conformance

As an interoperability framework, the FHIR specification provides mechanisms not only for describing the expected content of an exchange, but for validating that the resources involved in the exchange conform to the expected profiles.

The FHIR conformance framework supports the following types of conformance validation:

* Structure - Validation that data elements conform to structural requirements (i.e. resource and data element names)
* Cardinality - Validation that data elements are present in the expected cardinalities (e.g. 0..1, 1..1, 1..\* and others)
* Constraint - Validation that data elements satisfy expression-based constraints (e.g. if a name is present, a contact point must be as well)
* Terminology - Validation that codes used for data elements are from specific code systems and value sets
* Relationship - Validation that references between data elements conform to expected profiles.
* Capability - Description of capabilities such as search, profile support, terminology capabilities and operations.

These conformance capabilities provide a robust framework for the validation of healthcare exchange, and quality measurement use cases can make use of this framework to ensure data quality, consistency, and correctness. The QI-Core profiles, built from the harmonization of decision support and quality measurement data requirements, provide a basis for these use cases that can be expanded over time as new use cases evolve, and harmonization with other domains progresses.

Further, these profiles and implementation guidance can serve use cases for narrower exchange partners, enabling knowledge-sharing use cases within institutions, provider organizations, or other groups focused on particular specializations.

## Tooling

Because of the focus on implementation, the FHIR tooling ecosystem is both well-developed and well-supported. In addition, because of a strong emphasis on open standards and technologies, this capable tooling stack is largely open source, including:

* Publication tooling to support creation and maintenance of FHIR implementation guides
* Authoring tooling to support development of profiles, example instances and test cases
* Generation tooling to support simulation of large amounts of realistic clinical data
* Open source server implementations on multiple platforms and technologies
* Open source client implementations on multiple platforms and technologies
* Terminology tooling on multiple platforms and technologies

In the development of any data exchange capability, all these types of tooling are necessary to support proper authoring, distribution, interpretation, and implementation. Development of a de novo conceptual model would require building these layers of tooling, or at the very least finding a way to leverage existing tooling to support implementation.

## Agility

The FHIR publishing ecosystem has evolved over the past several years to a mature and stable technology stack, along with a broad and growing community of authoring expertise. This allows integration authors to focus on expressing their use cases, and lets stakeholders quickly see the expected results.

This agility comes with the associated challenge of versioning, in that new versions of implementation guides, as well as of the base specification itself, are released in relatively quick succession. This challenge is not inherent to FHIR, however, it is a constant feature of any changing system. What is critical is that the specifications for exchange have a mechanism to support and deal with evolution over time. Building on the shared experience of HL7 publishing, as well as the FHIR community, FHIR has a well-established and mature versioning model, both for the base specification, as well as for the implementation guides delivered on top of it.

## Reusability

Recognizing that quality measurement is part of a much broader quality improvement ecosystem, the use of consistent standards across those domains will make it easier to share content and services developed in different areas, such as decision support, research, population health management, public health reporting, registry reporting, and others. By using FHIR, support for many of these other domains are already part of, or actively being developed within, the FHIR ecosystem.

For example, the CDC Opioid Prescribing Support implementation guide developed shareable clinical logic for calculating Milligram Morphine Equivalent dose across a patient’s opioid-containing prescriptions, a complex calculation which requires significant development effort as well as ongoing maintenance to ensure correct calculation and up-to-date knowledge of opioid-containing drugs. Although this calculation logic was developed as part of a decision support implementation guide, measure developers were able to make use of this content directly to express measure population criteria.

As another example, the DaVinci Gaps-In-Care use case is exploring the use of logic developed for quality measurement to support defining care gaps for use in coordinated care-management. Using the same data element descriptions already developed for the measure and expressed within the QI-Core profiles, care-gaps reporting can be built and shared using the same underlying exchange specifications. As quality reporting data is submitted periodically throughout the measurement period, receivers can detect cases that do not meet the standard of care described for a measure and notify the reporter, or otherwise initiate corrective action.

## Implementability

FHIR directly supports quality measurement use cases through the Measure and MeasureReport resources. In addition, because these are also FHIR resources, there is no impedance mismatch between the measure specification and reporting containers and the data on which they operate.

This native representation means not only that systems that already support other types of FHIR resources have a lighter lift to read and process Measure and MeasureReport resources, but that because the container (MeasureReport) is also a FHIR resource, measure data can be packaged and exchanged using the same infrastructure.

In addition, as discussed in the Alignment and Tooling sections above, the FHIR focus on implementation, together with the availability of well-supported tooling further reduces implementation burden.

And finally, the FHIR Maturity model provides a clear indication of the level of testing and development the resources and profiles involved in an implementation have received, informing investment and risk assessment.

# Conclusion

This paper has made the case that using the FHIR ecosystem offers significant advantages to the quality measurement and reporting use case. From the breadth and depth of the information model, to the conformance-testable data exchange, to the agility and capability of the tooling, there are numerous reasons to use FHIR. That is not to say that FHIR is not without its challenges for meeting the quality measurement use case. In particular, the relatively low maturity of some of the key resources required to specify some quality measures presents a significant challenge. However, the healthcare industry has already adopted FHIR and is moving forward with surprising rapidity; and this trend is only reinforced by the recent ONC and CMS Final Rules[[3]](#footnote-3).

1. The FHIR Maturity model illustrates this commitment to implementation as a critical factor in the development of the specification: http://hl7.org/fhir/versions.html#maturity [↑](#footnote-ref-1)
2. The FHIR Credits page illustrates the breadth and depth of this community: http://hl7.org/fhir/credits.html [↑](#footnote-ref-2)
3. https://www.cms.gov/Regulations-and-Guidance/Guidance/Interoperability/index [↑](#footnote-ref-3)