Integrated Knowledge Management (IKM) Volume 6

Version 1 - Last Updated 11/28/2023

Table of Contents

List of Figures

List of Tables

List of Editorial Rules

Part I. ANF Ballot

Table of Contents

1. Why Analysis Normal Form? A Normal Form for Clinical Statements

A *clinical statement* is a definite and clear representation of a clinically-significant fact or situation was observed to exist or to have happened, or that a particular procedure was requested. A clinical statement can be expressed as a *narrative* that provides a written account that can be naturally read by humans, as well as a *normal form* which is a machine-processable representation of the statement's data as a standardized and encoded fundamental form. Today, clinical statements are often represented in unpredictable and denormalized forms, which makes reliable and safe decision support challenging, and reduces the quality of other types of data processing.

Healthcare organizations are striving to become *high reliability organizations (HROs)*, characterized by high levels of safety under inherently risky, technologically-complex, and demanding conditions. [1] Deployment of EHR systems is nearly ubiquitous in the US and there is increasing opportunity to leverage standards-based clinical statements to improve the health of the population (or citizenry) through quality measures, case reporting, and decision support. The ability to measure and improve outcomes relies on consistent, high-quality data that was aggregated from a variety of systems. Analysis using normal form allows HROs to derive added knowledge from data and reach high levels and safety. [2] A *standard* normal form can help replicate HROs across our industry.

In this document, we present background on other logical *HL7* and *ISO* clinical statement models, and focus on the need for—and logical specification of—an *Analysis Normal Form (ANF)*. ANF is a **normal form** intended to safely and reliably support data analysis that can be used to aggregate data created using any standard or non-standard input form or exchange mechanism. The ANF Reference Model, is a *logical model* that describes a standard normal form for clinical statements and it belongs to the *CIMI* library of logical models.

ANF is a model for clinical statements used in analysis that meet the following criteria: *Understandable, Reproducible, and Useful* (URU) [3][4]

- ***Understandable***. The content of an *ANF statement* can be processed by health IT systems and understood by most healthcare providers, without reference to private or inaccessible information.
- ***Reproducible***. Multiple users or systems apply the ANF to the same situations and source data with an equivalent result.
- ***Useful***. The ANF statement is fit-for-purpose—it has practical value for data analysis, in support of clinical decision support, research, and population health that requires information aggregated across health IT systems.

This document describes how information systems can improve patient safety and outcomes by increasing the precision of clinical information using a normal form to enhance and support quality data and analysis.

1.1. Motivation: Why Do We Need ANF?

Information systems record and manage clinical statements using a variety of standard or ad-hoc models. However, both treatment and analysis of clinical statements require consistency not only at the format level (e.g. *CDA*, FHIR, V2) but also the content model (i.e. an instance of an *ISO/TS 13972* DCM, CIMI model, etc.). [5] In most cases the data quality is the greatest obstacle to analysis, but even in the case of structured, semantically-clear information, inconsistency across sources of information raises obstacles to analysis. Analysis of aggregate information managed by health information networks poses the greatest

challenge today because a meaningful use of data for patient outcomes or research requires a common format, semantic clarity, and quality data.

Not only is there a potential for a lack of consistency with representing clinical statements with current *detailed clinical modeling* efforts, but there is also further variation in how the data are entered into information systems by end-users. This reality has a direct impact on patient safety if a clinical statement is recorded and displayed differently across the continuum of care. Clinicians author clinical statements and enter them into their organization's EHR systems where they are represented as some type of *"Clinical Input Form"* (CIF). This concept describes the representation of any natural language processing or data entry mechanism used by clinicians to record clinical statements. Vendors may compete on usability which may result in proprietary CIF data, or, clinical statements are based on standards-based models (e.g. CIMI, openEHR archetypes). For the purposes of this document, the type or usability of CIF data structures are not in scope. We assume that **any suitably encoded** clinical statement may be normalized.

Ideally, clinical information is modeled in a manner that is most efficient for use. This is a problem because there are many different use cases for clinical information with a wide range of requirements. There is no single model that can be the most efficient model for all the various use cases. Maximum efficiency for each use case necessitates that any particular clinical information be available in multiple modeled forms. These models, although different in form, semantically represent the same information, and are known as *isosemantic models*. Any particular detailed clinical model exists within a family of isosemantic siblings.

Clinical statements can be expressed and documented in many different ways in EHR systems, where clinical input forms provide different options to document the same clinical statement. These differences pose challenges for how the data are modeled and stored, and therefore have implications on data retrieval, data analysis, and accuracy of clinical analysis results.

1.1.1. Variation by Implementation: Clinical Input Forms

Clinicians enter clinical statements into their organization's EHR typically in a manner that we call here clinical input form (CIF), or the manner in which information is presented to the clinicians and how they enter the data, such as by constraining the information to allow only certain values to be entered - for instance, through a drop-down list, radio buttons, or breaking up large chunks of related information into smaller parts, or through natural language processing.

Let's consider the following example, represented below, in which data collected by an EHR combines information reported by devices with findings and interpretation:

1. A vital signs monitor transmits the systolic and diastolic blood pressure including date/time and the id of the device.

2. The nurse marks the measurement as "verified".

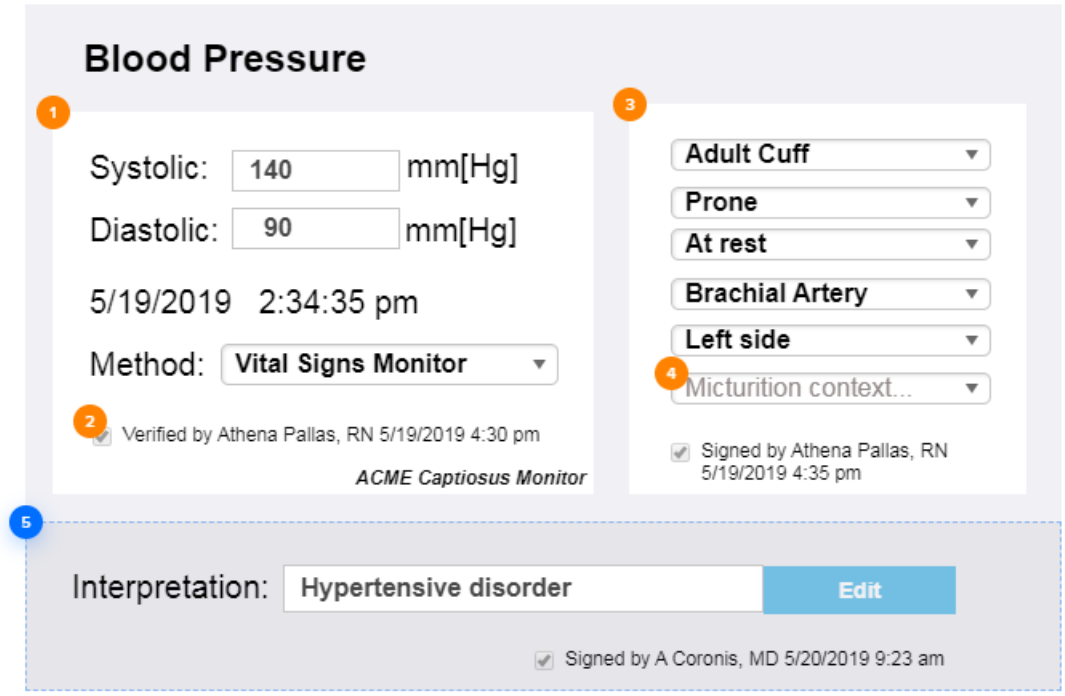
3. Next, the nurse documents how the measurement was performed:

   - using an adult cuff size
   - in prone position
   - brachial artery
   - on the left side
   - the micturition context is left empty/unknown[1]

4. Next, the physician adds an interpretation.

[1]Studies have shown that systolic blood pressure measurements could increase 10 to 15mmHg with a full bladder. Micturition, the process of emptying the bladder, is therefore a data element that can be recorded with some Clinical Input Forms. [6][7][8]

Figure 1.1. Blood Pressure Statement recorded by an EHR system



In this example the CIF provides the measurement information from the device to be verified by a nurse. The nurse adds annotations describing how the measurement was taken (at rest, prone) and the location (left brachial artery). The user may also fill in information about micturition, if known. A physician may interpret the measurement to be indicative of hypertension.

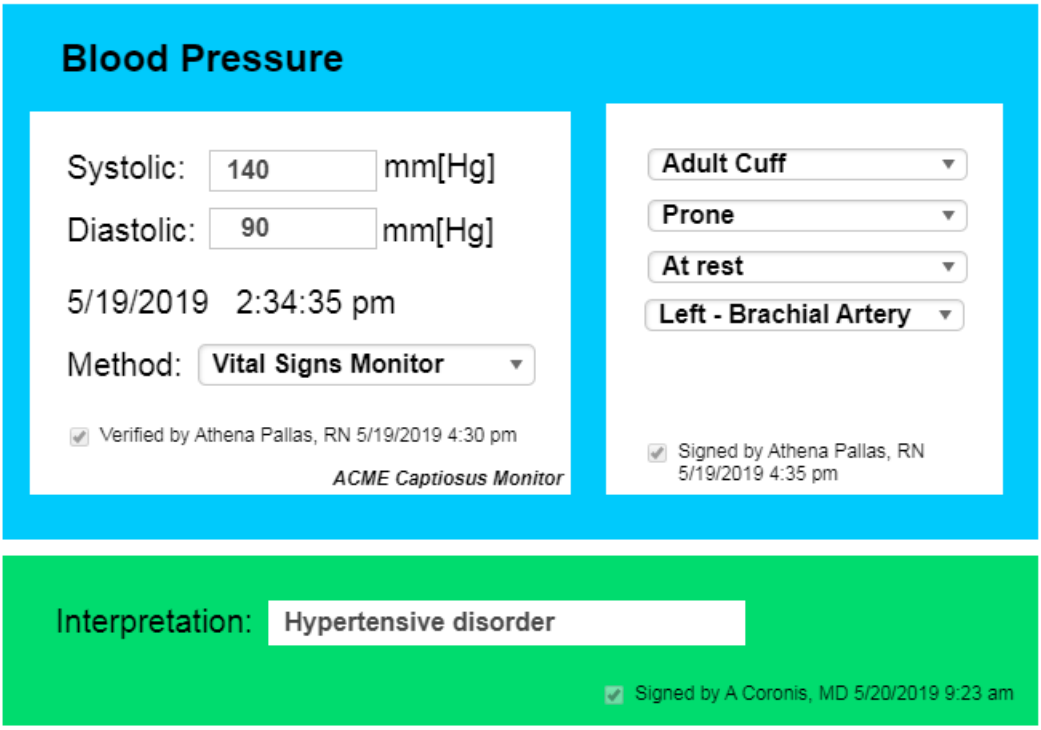
Another EHR system may capture or display a subset of information in CIFs about the blood pressure measurement—omitting "micturition context" and pre-coordinates site and laterality as:

- Right brachial artery
- Left brachial artery

The image below illustrates another distinct CIF in which the user interface captures a set of clinical statements related to Blood Pressure.

In the first case, the clinical input form has separate drop-down constraints to enter the artery and laterality as distinct concepts. In the alternative data entry form, the location and laterality are represented by a single, compound concept. This variation present in CIFs may also have implications on how the clinical statement is modeled, using different data elements to represent the same statement. Normalization to ANF eliminates the redundant structural variation and highlights the semantics of the topic.

Figure 1.2. Alternative Blood Pressure representation in a second EHR system



In this second CIF example, a similar system (or an alternative configuration of the same system) may support a different set of options to verify and record blood pressure measurement. This representation combines laterality and site and excludes details related to micturition.

1.2. Analysis Normal Form

Analysis Normal Form (ANF) is a logical model intended to represent a **normalized** view of aggregate clinical statements recorded during treatment for analysis, research, clinical decision support, and other purposes. ANF can be used to represent any isosemantic clinical statement irrespective of how the information was captured at its source (i.e. information systems or medical devices). ANF can be used in conjunction with other models intended to ensure that clinical information is structured and complete at the time of entry (e.g. CIMI models, ISO/TS 13972 Detailed Clinical Models) or exchanged among systems (e.g. HL7 CDA templates, *HL7 V2 message profiles*, FHIR profiles).

Clinicians, integrators, health IT developers, and researchers face different priorities, forcing trade-offs to be made that optimize data entry brevity at the cost of computability. ANF represents a collection of patterns and approaches to provide a predictable normal form to aggregate data sets across multiple systems. The more normalized a data set is, the simpler it will become to analyze, and errors will be reduced. In addition to improving analysis, ANF introduces the ability to compare statements with ease and no loss of semantic integrity.

1.2.1. Objectives and Purpose of ANF

ANF's purpose is to introduce standards-based, normalized representation of clinical statements from heterogeneous sources using an objective ***measure*** to help evaluate the result, presence, and magnitude of a specific finding, request or observation. ANF requires an ability to classify the ***topic*** of a statement using

standard terminology expressions. ANF defines responsibility for different representational aspects of input data along well-defined compositional layers (see Separation of Concerns). In practice, information systems may create normal data natively or transform other representations of clinical statements (e.g. Consolidated-CDA templates, FHIR profiles) to normal form (i.e., ANF).

Overall, ANF allows healthcare enterprises to normalize information aggregated across multiple sources to better support a set of analysis. ANF enhances the ability to analyze and compare clinical statements aggregated across systems and organizations and provides a logical model to:

- Specify a common form for clinical statements extracted from EHR systems and FHIR.
- Provide a common analysis form to data exchange paradigms (e.g. HL7 messages, FHIR and CDA).
- Enhance clinical data for use in *Clinical Decision Support* Systems, Clinical Quality Measures and National Registries, Healthcare Guidelines and Protocols, and Epidemiological Research.

1.2.2. Assumptions for ANF

ANF provides a precise statement specification that is comparable and sharable between multiple care providers, health enterprises, and standards-based Healthcare Information Technology (HIT) systems. ANF does not define the terminology specification but relies on *terminology knowledge* to specify the meaning of clinical statements. ANF-based data may use single codes, as well as any legal terminology expression defined within the terminology layer of the architecture.

ANF supports *pre-coordinated* and *post-coordinated* terminology expressions to provide greater content coverage than can be achieved by relying only on pre-coordinated concepts. Post-coordinated compositional terminologies are more expressive and can achieve better analysis than can be achieved by relying only on pre-coordinated concepts.

Successful analysis requires appropriate data quality necessary for systems to define a precise topic, type, and clear measure or *result* of what was observed, requested, or assessed during treatment. ANF can be applied to any input data and any formalism as long as the data semantics and terminology are sufficiently precise to define the elements mandatory for analysis.

1.2.3. Approach - Architectural Separation of Concerns

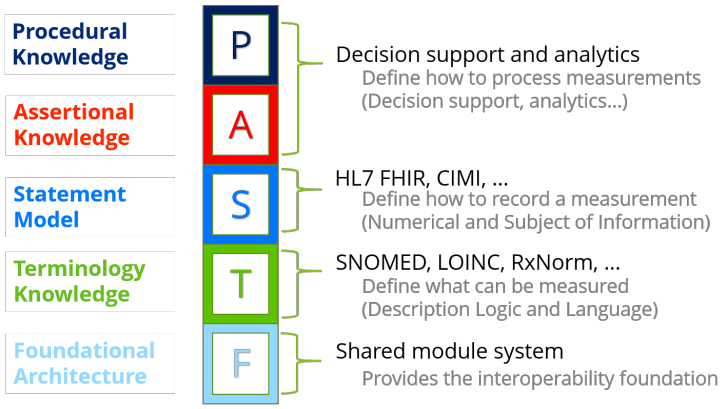
Increased reliance on computerized health records, including Electronic Health Records Systems, requires standardized medical terminology to encode health information consistently across systems and enterprises. Clinicians require not only objective quantitative measurements (e.g. 90 beats per minute for a patient's pulse) but also contextual or procedural context (e.g. pulse oximetry, manual) about past observations or requests for future interventions. While two quantitative measurements may be the same, the procedural information could indicate meaningful semantic differences and lead to different clinical interpretation and treatment. As information is exchanged across systems, the solution requires a common understanding of data and a method to support knowledge-representation and clinical decision rules based on common terminology and statements. Each component must address an aspect and, together they need to address the requirements of clinicians. Current HL7 standard implementations rely on profiles and templates to disambiguate statement and terminology, and provide sufficient precision for transactions, documents, and standards-based APIs. Therefore the architectural approach described here would be applicable to standards organizations developing interoperability-enterprise, and project-specific implementations in equal measure.

Functional decomposition—often referred to as a *Separation of Concerns* (SoC)—across components or sections with a specific purpose is a foundational design principle for complex system architecture. Enabling a SoC allows a complete system to be subdivided into distinct sections or components with well-defined functionality and dependencies. If successful, this approach allows individual sections to be able

to be *reused*, as well as designed, implemented, and updated *independently* to address emerging *requirements*. This is especially useful and important in a medical context given how many different health information and clinical terminology projects are ongoing at any given time. Efforts are often uncoordinated and led by disparate and unrelated standards development organizations. In these cases, SoC allows teams to work independently, in coordination with each other, and reuse the resulting artifacts.

Figure 1.3, "Separation of Concerns: Knowledge Architecture" shows how a layered knowledge architecture can enable a separation of concerns.

Figure 1.3. Separation of Concerns: Knowledge Architecture



Separation of concerns is an architectural design principle, whereby a system is divided into distinct sections, such that each section can address separate concerns. In this case, each architectural layer may build upon artifacts from lower layers.

Foundational Architecture – The *Foundational Architecture* of the Knowledge Architecture provides the common elements of interoperability such as object identity, versioning, modularity, and knowledge representation. It includes a) the foundation and building blocks of the common model; (b) how the repeatable transformation process of disparate standards into the common model promotes interoperability with other environments; and (c) how the modules of the architecture are tightly version controlled over time.

Terminology Knowledge – The Terminology Knowledge layer is responsible for structured sets of medical terms and codes that *define* concepts of interest, including descriptions, dialects, language, and semantic hierarchy. SNOMED CT, LOINC, and RxNorm are part of this layer. It defines what valid codes or expressions may be used by higher level layers.

Statement Model – The Statement Model layer is responsible for defining how data elements are combined to create a statement. ANF Reference Model belongs in this layer. Other standards-based clinical statements are discussed later in this chapter. This layer reuses the artifacts defined in the Terminology Knowledge layer.

Assertional Knowledge – The Assertional Knowledge layer makes use of the Terminology Knowledge layer concepts to specify *non-defining* facts that may be used by procedural knowledge algorithms. An example of such a fact might be that "thiazide diuretics treat hypertension." Assertional Knowledge may indicate what symptoms may be associated with a disorder.

Procedural Knowledge – Procedural knowledge, also known as imperative knowledge, is the knowledge exercised in the performance of some task, such as determining a hypertension treatment plan by analyzing a combination of a patient's ANF statements, and the available assertional knowledge. The procedural knowledge is responsible for information about standard ways to carry out specific procedures as well as other procedural guidelines, e.g. treatment protocols for diseases and order sets focused on particular patient situations. Procedural knowledge, together with assertional knowledge, enables clinical decision

support, quality measurement, and supports patient safety. This layer relies on the architectural foundation and terminology layers, incorporates the statement model for information retrieval, and uses the assertional knowledge. Procedural knowledge artifacts may include clinical alert rules, reminders, etc. that trigger actions or recommend interventions.

Examining a clinical procedure for controlling hypertension illustrates each of the layers of the informatics architectural separation of concerns.

- At the Terminology Knowledge layer, there may be various codes and terms from disparate source terminologies to define a concept (e.g. hypertension). Ideally, these overlapping codes and terms would be oriented to the same parent concept during the transformation and integration process at the Foundational Architecture layer (e.g., *Solor*).
- The Statement Model layer enables representation of blood pressure measurement values (e.g., systolic BP = 140 mmHg) or the categorical data (e.g., pregnancy induced hypertension vs. renal hypertension) within a standard data structure to facilitate information exchange or retrieval, such as within a standards-based clinical statement (i.e. CIMI, CDA, FHIR, ANF, etc.).
- The Assertional Knowledge layer represents non-procedural statements, or facts, such as "Stage 2 high blood pressure is over 140 systolic or 90 diastolic," or that beta-blockers and ACE inhibitors may be used to treat hypertension, or that beta-blockers are contraindicated in patients with a diagnosis of reactive airway disease.
- Finally, the Procedural Knowledge layer provides algorithms to analyze ANF statements about a patient, in combination with the Assertional Knowledge, to recommend a treatment protocol for different kinds of hypertension, including the considerations of, e.g. patient age, comorbidities etc., which can be generated by an electronic clinical decision support system (Statement + Assertional layers). This layer adds support for workflow and conditional logic (i.e. if-then-else).

A clear separation of concerns enables the isosemantic transformation of standards-based clinical statements to normal form in the Statement Model layer by decoupling structure from semantics and workflow.

HL7 relies on implementation guides (for V2, CDA, and FHIR) to add sufficient terminology knowledge to standards-based clinical statements. Vocabulary constraints documented as profiles or templates are the mechanism to create interoperable implementation guides from health IT standards. Only after the Terminology Knowledge is fully defined, the standards-based statements can be used to support business and workflow decision points consistent with the Assertional and Procedural layers described above.

1.3. Background: HL7 Clinical Statement Standards

Clinical statement standardization has been a long-standing concern for HL7 and reuse of these content models across paradigms (e.g. messages, documents, services). Standardization has relied on model-driven approaches requiring a separation of concerns along with conceptual, logical, and implementation perspectives.

HL7 Service-Aware Interoperability Framework (SAIF) organizes HL7 standards along three perspectives (i.e. conceptual, logical, and implementable).

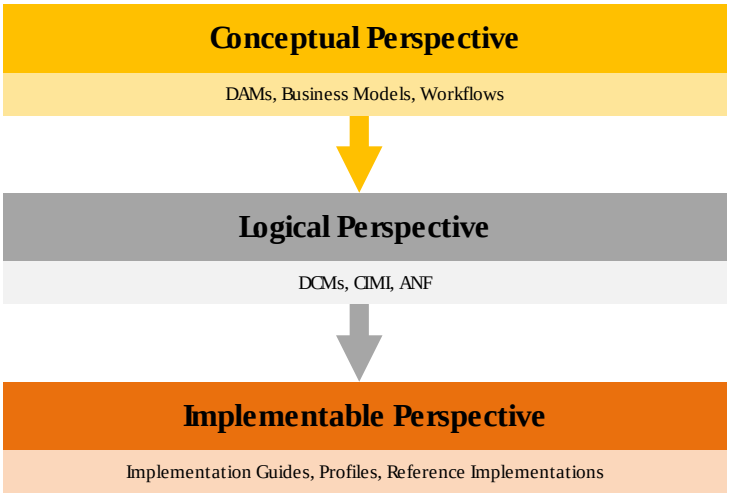
1.3.1. HL7 Service-Aware Interoperability Framework (SAIF)

To augment the HL7 Version 2 and Version 3 representations, HL7 introduced an architecture to allow for a clear separation of concerns among interoperability models and specifications from the abstract or

conceptual to the most precise, implementable, and testable that ensures semantic interoperability. This architecture is the HL7 Service-Aware Interoperability Framework Canonical Definition (SAIF-CD).[9]

The SAIF-CD specification [9] defines three SAIF Perspectives: Conceptual, Logical, and Implementable. These perspectives are not formally equivalent with *Object Management Group's (OMG)* levels-of-abstraction in *Model-Driven Architecture (MDA)* even though it reuses the same derivation. Therefore, the Implementable Perspective is derived from the Logical Perspective and the Logical Perspective is derived from the Conceptual Perspective. This approach ensures that any implementable artifacts (i.e. service specifications, implementation guides) are traceable to business/clinical requirements and logical models of knowledge.

Figure 1.4. Model Derivation based on SAIF-CD



Like other CIMI models, ANF is a Logical Model that may be used to create implementation specifications.

However, the SAIF Conceptual Perspective is not completely equivalent to the MDA concept of *Computationally Independent Model (CIM)*, the Logical Perspective is not equivalent to the MDA *Platform Independent Model (PIM)*, nor is the Implementable Perspective equivalent to the MDA *Platform Specific Model* although this Perspective is the SAIF Perspective that most closely aligns with an MDA analogue.

1.3.1.1. Conceptual Perspective

These artifacts are most commonly focused on the "Problem-Space" rather than the "Solution-Space," and contain explicit, unambiguous descriptions of the various dimensions of the component (e.g. clinical statement) or system being specified.

A fully-specified Conceptual Perspective thus should be both readable and traceable by Domain Experts and Subject Matter Experts and rigorous enough to serve as input into the development in the Logical Perspective.

In HL7, the Conceptual Perspective is represented by *Domain Analysis Models (DAMs)* and business models that represent stakeholder requirements analyzed by subject matter and domain experts. This perspective precedes the development of either logical or implementable artifacts and it is key to successful testing of implementations.

1.3.1.2. Logical Perspective

Artifacts in the Logical Perspective represent traceable translations of Conceptual-level artifacts into a form and format, usable by and useful to architects and "inward-facing analysts." Also included are ad-

ditional specification materials required by architects preparing artifacts for consumption by developers. The Logical Perspective contains platform-independent artifacts.

There are no definite boundaries between the Logical and Implementable Perspectives. Therefore, it is important for organizations such as HL7 to standardize logical models used to generate/create implementable artifacts (i.e. implementation guides, profiles, and templates). *CIMI Clinical Statements*, ISO/TS 13972 Detailed Clinical Models, and ANF statements all belong in this perspective.

1.3.1.3. Implementable Perspective

Artifacts in the Implementable Perspective are typically defined by developers or standards implementers, often through discussion with software designers, architects, or system integrators. Note that the artifacts in the Implementable Perspective are not actual implementations, but rather implementable in a number of implementation instances. Thus, all the necessary technical bindings, including data types, value sets, class libraries, and interface specifications are part of the Implementable Perspective. FHIR implementation requires a combination of profiles and test cases to ensure that implementations meet the requirements used to derive the conceptual and logical models.

1.3.2. HL7 Version 3 Clinical Statement

Starting with HL7 Version 3 [10] the minimum requirements for the interoperable clinical statement are:

Table 1.1. HL7 V3 Clinical Statement Definition

"Clinical Statement for the care of patients (persons, animals and other entities) is: *An expression of a discrete item of clinical, clinically-related or public health information that is recorded because of its relevance to the care of a patient or other entities. Clinical or public health information can be expressed with different levels of granularity and therefore the extent and detail conveyed in a single statement may vary. To be regarded as a Clinical Statement, a concept must be associated with a patient or other entity in a manner which makes clear:* • *Its temporal context* • *Its relationship to the entity or entities* *In the case of an observation, its mood and presence, absence or value* *In the case of a procedure, its mood and status* *This clarity may be achieved by:* • *Explicit representation; or,* • *Implicit application of defaults ONLY where explicitly modeled rules state the appropriate defaults."*

The V3 Clinical Statement Model is applied across CDA implementation guides including the US-Realm Consolidated CDA (C-CDA) to represent CDA document entries. A V3 Clinical Statement Model is a *polymorphic model*: it can represent observations, procedures, encounters, public health reports, supply, medications, exposure, and derivations of clinical acts. The V3 Clinical Statement model provides a Statement Model with partial Terminology constraints. For example, Clinical Statements in a CDA document section need to be constrained to add the precision needed to support the Terminology Knowledge layer. CDA entry and sub-entry templates can be used to create precise implementations of the V3 Clinical Statement model for a specific type of clinical statement (e.g. Procedure Activity, Problem Observation) sharing a common statement model but different terminology and usage constraints. The US-Realm C-CDA specification consists of a set of templates that constrain the document, sections, and entries used in each section.

1.3.3. CIMI Statements

The *Clinical Information Modeling Initiative (CIMI)* is defining a library of logical *clinical information models* using a common modeling formalism. CIMI intends to improve the interoperability of healthcare information systems through shared detailed clinical information models that can be used to generate platform-specific model specifications such as FHIR profiles, CDA templates, OpenEHR Archetypes, ISO 13606 Archetypes, ISO/TS 13972 DCMs. CIMI models are grouped into semantically equivalent (or 'isosemantic') families of detailed clinical models, which capture the same clinical meaning using different combinations of *pre-* and *post-coordinated concepts* and corresponding information model structure. The central focus of the CIMI Reference Model is the CIMI Clinical Statement. A CIMI Clinical Statement represents structured electronic communication made about a patient typically documented as an 'entry' in the patient record.

Unlike the V3 Clinical Statement Model applied in C-CDA, CIMI models are designed with Terminology Knowledge and provide a separate model for each type of statement, organized into a comprehensive library.

For reader convenience, CIMI clinical statements are further explained in an appendix of this document. (See Current CIMI Modeling Efforts).

1.3.4. Related ISO Standards

ANF is intended for projects that aggregate clinical statements from a variety of sources, independent of formalism or approach used by the source system. It normalizes approaches and methodologies in use across the industry and provide a uniform representation of data to enable analysis in a platform-independent view using context-free languages. ANF underscores that both treatment and analysis of clinical statements require consistency not only at the format level (e.g. CDA, FHIR, V2) but also the content model (i.e. an instance of an ISO/TS 13972 DCM, CIMI model, etc.). Context-free languages offer a high level of expressivity and formalization, thus enabling the representation of any real-world artifact. ANF can be applied to real world business systems using platform-specific representations (e.g. FHIR profiles and resources, database schema definitions) Those representations may be derived from the ANF Reference Model. Like other standards for defining and managing clinical statements such as ISO 13606 EHR Communication or ISO 13972 Clinical Information Models, ANF may be extended by the ISO 23903 Interoperability and Integration Reference Architecture to enable the justification of correctness and consistency of the content models and their relationships.

Different use cases usually represent different contexts, frequently not reflected in the provided abstraction of the models. However, the context impacts the semantics and therefore the ANF logical model allows clinical statements to be associated/related to other statements to describe complex clinical data.

1.4. About this Document

This document describes how information systems can improve patient safety and outcomes by increasing the precision of clinical information using a normal form to enhance and support quality data and analysis. In the subsequent chapter we will provide a deep dive into the building blocks and constructs for ANF, in a chapter containing the ANF Reference Model and illustrative examples of ANF modeling.

Subsequently, we will outline how the various building blocks and attributes work together to create ANF Clinical Statements. We then provide the ANF Modeling Methodology, including a list of modeling principles and rules. Next, we discuss how clinical statements can be transformed and normalized into ANF Clinical Statements. Finally, we discuss the implications of ANF on data quality, clinical decision support, and ultimately, patient safety and outcomes. In the appendices, we explore current CIMI modeling efforts

including illustrative examples for modeling CIMI clinical statements. We also compare and contrast ANF Clinical Statements and CIMI Clinical Statements in an appendix.

Note

SNOMED CT is used as a representative example of a terminology system for the coded data elements in the ANF Reference Model. This ballot is focused on defining a Statement Model, not the underlying Terminology Knowledge layer described in the Knowledge Architecture. While the SNOMED CT examples are based on actual SNOMED CT definitions that are part of the SNOMED CT distribution, we recognize that there are inconsistencies within SNOMED CT that allow redundant representations. A first step in addressing the potential for redundant representation in the Terminology Layer is to define a separation of concerns between the Terminology and Statement layers to eliminate redundant representations between layers. Subsequent efforts to improve the quality of the Terminology Layer can then be done independent of the Statement layer.

2. Building Blocks: ANF Reference Model

The *ANF* Reference Model is a logical information model describing the format of a normalized clinical statement that may have originated from an information system data store, a standard-based message (e.g. HL7 Version 2), a standard-document (e.g. *HL7 CDA*), a standard-based resource (e.g. *HL7 FHIR*), or an instance of a *CIMI model* (e.g. FHIR-based profile, *openEHR* archetype).

The ANF Reference Model describes the **normal form** proposed by ANF. Along with the editorial rules the ANF Reference Model describes how to reduce data redundancy to support analysis of aggregated clinical statements. A clinical statement expressed in the ANF Reference Model is in Analysis Normal Form if and only if it conforms to all the Editorial Rules defined by this specification.

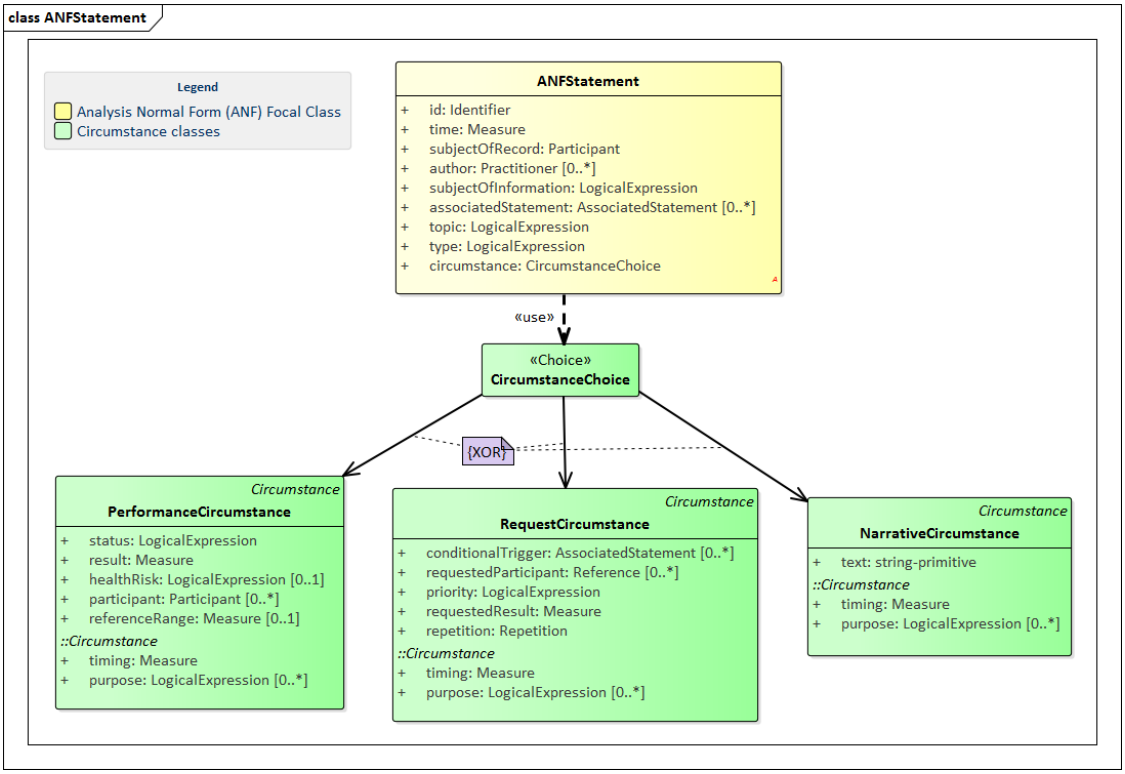
Similar to database schemas, clinical statements often include redundant information about the topic, result, and circumstances of an action performed or requested. Often clinical statements include multiple ways to represent a request or performance. ANF provides a target normal form along with rules/guidelines for normalization to make data available for analysis.

2.1. Model Representation

The ***ANF Reference Model*** is a logical model described herein using the *Object Management Group (OMG) Unified Modeling Language (UML)* 2.0 notation to describe the structure of normalized *clinical statements* for computational analysis. This *logical model* may be implemented using any programming language, database technology, or interoperability specification (e.g. FHIR) suitable for analysis. ANF is intended to normalize approaches and methodologies in use across the industry and provide a uniform representation of data to enable analysis.

The following diagram describes the logical structure of a clinical statement that conforms to the Analysis Normal Form specification. At a high-level an *ANF statement* defines the topic (**WHAT** happened, was observed, requested, measured, asserted, etc.) and under what circumstances-(**HOW**, **WHY, WHEN, and with what RESULT**).

Figure 2.1. ANFStatement Structure



2.1.1. ANFStatement

This is the main class of the ANF reference model and it describes the structure of a normalized clinical statement that complies with ANF.

Attribute	Multiplicity	Notes
id Identifier	[1...1]	Unique identifier of the statement.
time Measure	[1...1]	This data element describes when the statement was documented. Is it's expressed as a **Measure**. For example the date of **2019-07-09T00:12:31+00:00** would be represented as Unix Epoch time as **1562631151** seconds: • interval.lowerBound = 1562631151 • interval.includeLowerBound = true • interval.upperBound = 1562631151 • interval.includeLowerBound = true • semantic = Seconds, Unix Epoch Time The ANFStatement separates the timing related to documenting a statement vs, the timing of the phenomenon

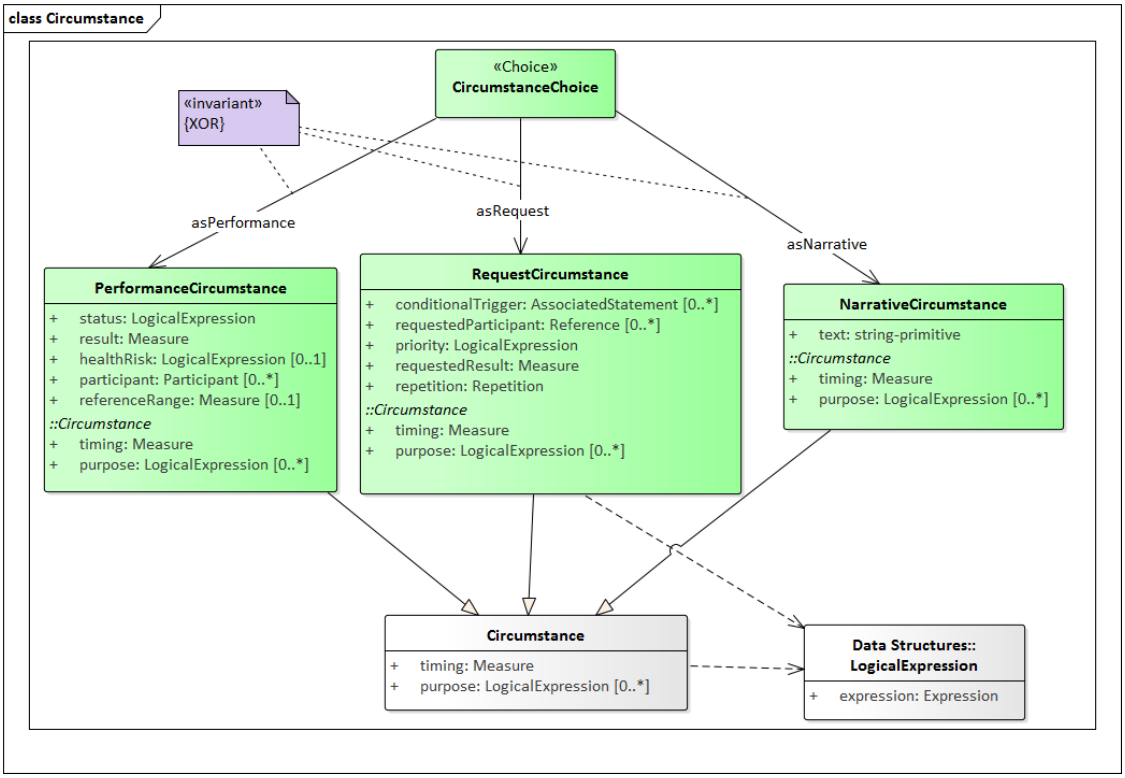
Attribute	Multiplicity	Notes
		that the statement is describing. This data element specifies when the statement was recorded/asserted.
subjectOfRecord Participant	[1...1]	A patient's clinical record will contain many statements. The subjectOfRecord is a reference to the patient clinical record in which this statement is contained.
author Practitioner	[0...*]	Options list of identified authoring practitioners.
subjectOfInformation LogicalExpression	[1...1]	A logical expression describing the subject of the statement; it's used to express **WHO** the clinical statement is about. A patient's clinical record may contain statements not only about the patient, but also statements about children, relatives and donors. Thus, some possible values for subjectOfInformation would include codes for 'subject of record' (the patient), 'family member', or 'donor'. The majority of statements will have a subjectOfInformation with a value of 'subject of record', since most statements in a patient record will be about the patient. The subjectOfInformation is used to represent who the statement is about. This is normally the patient unless explicitly stated otherwise.
associatedStatement AssociatedStatement	[0...*]	An ANF statement associated to the current ANFStatement. If the topic is a laboratory result panel, each association would point to another statement which is a laboratory result. The semantic of the associated statement may be: • a precondition • an interpretation • an component (observation)
topic LogicalExpression	[1...1]	This data element is an expression of **WHAT** is being requested or what was performed. For both ANFStatement types (request or performance) a *pre-coordinated* or *post-coordinated* “procedure” concept as a logical expression is required to sufficiently capture the action, which is either requested or performed. The topic is the central component of clinical statements. The following are proposed principles for the topic of an ANFStatement. **Principle 1:** The topic defines the action that is being requested, measured, or observed. **Principle 2:** The topic has to be able to exist on its own and still retain original intent and clarity of meaning.

Attribute	Multiplicity	Notes
		Principle 3: Each clinical statement may only have one topic [but the topic is a comprehensive expression].
type LogicalExpression	[1...1]	This data element distinguishes between a performance ('**performed**') and a request ('**requested**'). Performances may be observational performances, e.g. the observation of a clinical finding or disorder being present or absent. They can also be a procedure or intervention which has been performed on the subject of record in the past, e.g. “a procedure using a 12-lead electrocardiogram”. Performances can – but do not have to – include quantitative or qualitative results, e.g. “3 dot blot hemorrhages” or “Hepatitis A antibody positive”.
circumstance CircumstanceChoice	[1...1]	A choice of circumstance appropriate to the type of clinical statement.

2.1.2. Circumstance

Circumstances can describe **HOW**, **WHY, WHEN, and with what RESULT** a requested or performed action will be or was carried out. ANF promotes a normalized representation of observation or intervention results where all results are reduced to a "measure". This approach reduces data retrieval difficulties by eliminating the potential for multiple differing representations of the same clinical statement. For example, with coded results there are multiple potential methods to represent eye color that complicate data retrieval. The Topic could be a Finding refined by an Observable (Iris finding->Interprets = Color of iris) or a Finding with no refinement (Finding of color of iris). In both of these cases the Result would be a qualifier of Blue color. The ANF Statement would represent Eye color using the Blue iris Finding as the Topic and the Result would be Present, represented as interval.lowerBound =1, interval.upperBound=INF.

Figure 2.2. Circumstance



2.1.2.1. CircumstanceChoice

This class provides an exclusive choice of circumstances that may be chosen when an ANFStatement is instantiated:

- PerformanceCircumstance
- RequestCircumstance
- NarrativeCircumstance

2.1.2.2. Circumstance

This abstract class is used to describe the default data needed describe any circumstances associated with a clinical statement.

Attribute	Multiplicity	Notes
timing Measure	[1...1]	Timing is used to capture a time or time range for: • Requests for action at a future time • Performance of action, which has taken place in the past (including “History of X….) • When an action was supposed to be performed • When an observed finding or disorder was present or absent

Attribute	Multiplicity	Notes
		• When a procedure was performed
purpose LogicalExpression	[0...*]	This data element describes in a post-coordinated expression the reason for a performance or request. For example in the case of a procedure, the purpose may be either: • 386053000 |Evaluation procedure (procedure)| • 277132007 |Therapeutic procedure (procedure)| The procedure is then refined by post-coordinating with a "363702006 |Has focus (attribute)|" attribute and identifying a finding/disorder or procedure concept as the value for the attribute.

2.1.2.3. RequestCircumstance

This class further specifies **HOW** a requested action is to be performed, e.g. how often or how long.

A Request for Action clinical statement describes a request made by a clinician. Most of the times, but not always, the object of the request (e.g., lab test, medication order) will be fulfilled by someone other than the clinician (e.g., lab technician, pharmacist) making the request. All detailed information about the request will be documented in this clinical statement, such as patient must fast for 12 hours before having a lipids blood test.

Examples:

- Request for Rheumatoid factor 1 time routine
- Request for X-ray chest to evaluate for heart failure
- Cardiology referral
- Ribavirin 200 mg capsule oral, take 2 capsules every morning
- Advised to participate in tobacco cessation counseling once a week.

Attribute	Multiplicity	Notes
conditionalTrigger AssociatedStatement	[0...*]	This data element is used to represent a condition, or set of conditions that must exist in order for Request to be executed. For example, Ibuprofen 400 mg tablet oral every 6 hours as needed for back pain, the use of Ibuprofen is conditional on the presence of back pain.
requestedParticipant Reference	[0...*]	This data element is an optional list of either specific persons or roles who perform an action, assist in performing an action or are targets of an action.
priority LogicalExpression	[1...1]	This data element specifies the priority with which a requested action has to be carried out, e.g. "routine" or "stat". By default a Request will be considered "routine" unless otherwise specified.

Attribute	Multiplicity	Notes
requestedResult Measure	[1...1]	This data element specifies the measurable result; it may specify that something must be completed (e.g. how many sessions of counseling, how many refills, etc.) are requested or that something be done.
repetition Repetition	[1...1]	This data element describes when an action is requested for more than a single occurrence using the Measure data structure: • When the repeated action should begin (periodStart), e.g. NOW • How long the repetitions should persist (periodDuration), e.g. for 3 weeks • How often the action should occur (eventFrequency), e.g. 3 times per week • How long between actions (eventSeparation), e.g. for 2 weeks • How long every action should last (eventDuration), e.g. for 5 minutes

2.1.2.4. PerformanceCircumstance

This class describes the circumstances associated with a statement. It is used when an action or observation is performed and it specifies the result of intervention using both measure and a coded status .

For example, "*Insulin placed on hold 24 hours prior to catheterization*" would have a status of "*On hold*". A typical, successfully completed procedure would have a status of "*Completed*".

Attribute	Multiplicity	Notes
status LogicalExpression	[1...1]	This is a coded value representing the current status of the intervention (e.g. "completed"). This data element is not intended as a substitute for workflow specification.
result Measure	[1...1]	Intervention result as a measure.
healthRisk LogicalExpression	[0...1]	This optional data element is used to flag a result with coded values to describe the health risk associated with result of the ANF statement (e.g. 'low', 'normal', high', 'critically low', or 'critically high'). **Note**: this data element is not equivalent to an **interpretation**. Interpretations of clinical statements are represented as related ANF statements (e.g. "hypertensive disorder" statements - interpretation of a 'high' systolic blood pressure).
referenceRange Measure	[0...1]	This optional data element is the interval of values that are normal for the observation/finding described by the "topic" for this "subject". It refers to reference for the patient/subject with these conditions.
participant Participant	[0...*]	This optional data element identifies the practitioner(s) responsible for the results reported.

2.1.2.5. NarrativeCircumstance

This class is used to describe the circumstances of a clinical statement using natural language/text rather than a structure.

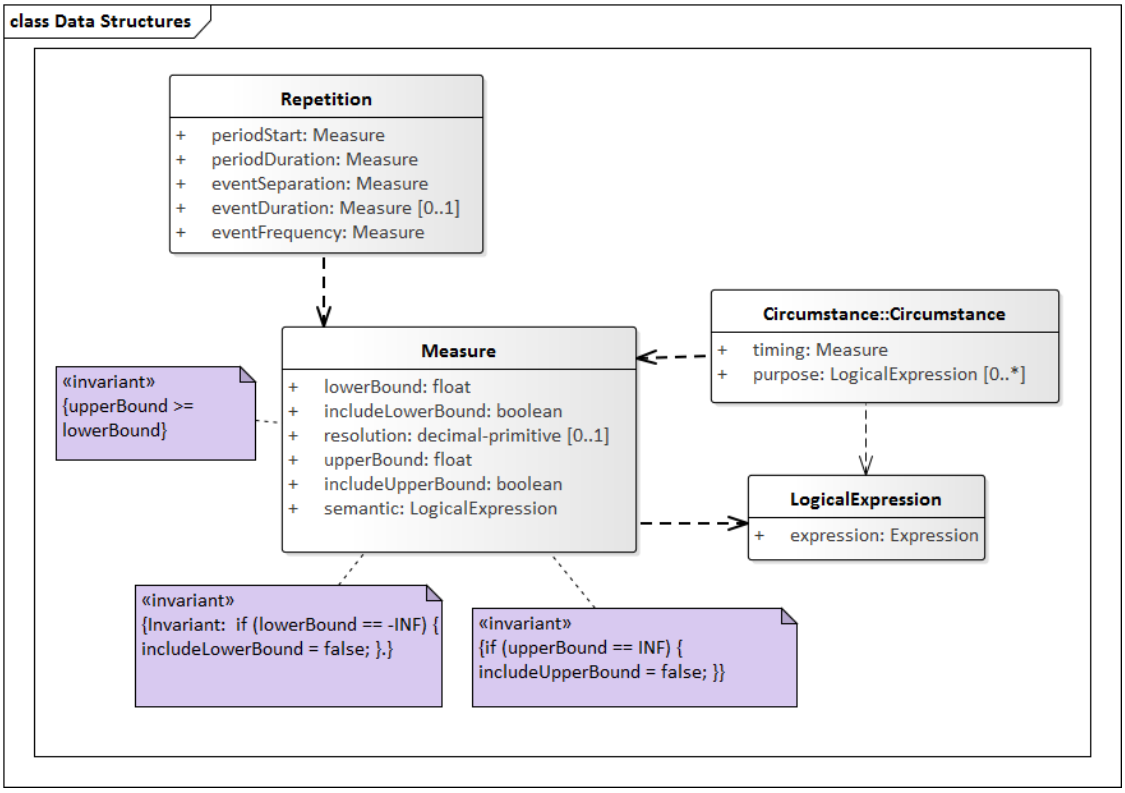
This class may be used to specify either a performance or request circumstance.

	Multiplicity	Notes
text string-primitive	[1...1]	Text description of circumstances.

2.1.3. Data Structures

The following are data structures used to represent an ANFStatement. This section describes the data structures specific to ANF. This model references a set of logical structures to represent unique identifiers (i.e. Identifier, Expression) and primitive types (boolean, float).

Figure 2.3. Data Structures



2.1.3.1. Measure

This class captures measurable elements of clinical statements, e.g. the results of test procedures, time periods, frequencies of repetitions for procedures or medication administrations. The measure formally represents a numeric interval between two non-negative real numbers with a semantic and precision/resolution. The interval can be open or closed depending on whether the upper and lower bounds are included in the measure interval.

The measure provides a single way to represent both "presence" or "absence" values and numeric values for a phenomenon. In general, the interval value represents the numeric range within which the observed

value of a phenomenon occurs. Note that this formalism allows both exact values and ranges of values to be expressed. In the case that the beginning and end points of an interval are the same value, the meaning is that the value of the phenomenon is exactly that value.

- [10, 10] : means the value is exactly **10** because the *lowerBound* and *upperBound* are both **10**;

In the special case that the beginning of the interval is a number, **n**, and the end point is **INF** (infinity), the meaning is that the value of the phenomenon is > **n** or **>= n**, depending on whether the interval is open or closed.

- (0,INF) : > 0 ; (greater than 0)
- [10,INF) : >= 10 (greater than or equal to 10)
- (-INF, 10): <10 (less than 10)

The interval value may also represents whether a phenomenon is "**present**", "**absent**", or "**indeterminate**". Specifically, any interval value that includes only numbers that are greater than zero also denotes the value "present".

Any interval value that includes only the number 0, itself, denotes the value "absent". Any interval value that includes both the number 0 and at least one number > 0 denotes the values "indeterminate". Lastly, there are two interval values that explicitly denote "present" and "absent", respectively. These values may be assigned to phenomena that would not otherwise take on a numeric value (such as "nausea"):

- Nausea value = (0,INF) : **present** (an open interval that excludes zero)
- Nausea value = [0, INF): **indeterminate** (an open interval that either both presence or absence)
- Nausea value = [0,0] : **absent** (exactly zero)

The numeric attributes of this class are of type "**float**" to support both positive and negative values that conform to IEEE 754 standard for Floating Point Numbers.

Note

For Java implementations, a **float** number uses 32 bits to represent the sign, exponent, and mantissa consistent with IEEE 754:1985. The values **+infinity** and **-infinity** are denoted with an exponent of all ones and a mantissa of all zeros. The sign bit distinguishes between negative infinity and positive infinity.

Attribute	Multiplicity	Notes
lowerBound float	[1...1]	It specifies the lower bound of a measurable element. This can be the lower bound of a range: • For the “Tumor greater than 1 cm but less than 4 cm” the lower bound is 1. • For a test result, which is not a range, lower and upper bound are the same. Example: systolic blood pressure 110 mmHg. The lower and upper bound are both 110 mmHg. • For an unbound measure, the lowerBound is -INF(negative infinity) and *includeLowerBound* is "false"
includeLowerBound boolean	[1...1]	It states whether the lower bound in the interval is included in the interval. The inclusion or exclusion of lower bound is needed to

Attribute	Multiplicity	Notes
		express measurable elements which include relative properties, such as “greater than”, “less than” and others. **Example**: “Persistent cough for more than 10 days”. If a lower bound of “10” is chosen, it would not be included, because the example states: more than 10 days. Choosing “11” would require it to include the lower bound. If "true" the lower bound is part of the interval. **Invariant**: if (*lowerBound* == - INF) { *includeLowerBound* = false }.
resolution decimal-primitive	[0...1]	It defines the possible or allowed increments in which the measured “thing” can be counted. In the example of the systolic blood pressure of 120 mmHg, the resolution is “1”, because the blood pressure measurement result can be counted in 1 mmHg increments. The Resolution is not always defined or known. Example: a clinical statement like “History of breast cancer” implies an undefined amount of time in the past and it is not stated if it is years, months, etc.
upperBound float	[1...1]	It represents the upper bound of a measurable element. This can be the upper boundary of a range: For the “Tumor greater than 1 cm but less than 4 cm” the upper bound is 4. In cases, where the measurable element does not represent a range, upper and lower bound have the same value. **Invariant**: *upperBound >= lowerBound*.
includeUpperBound boolean	[1...1]	It states whether the upper bound in the interval is included in the interval. Similar to lower bound, where the measurable element has relative properties, the same rules apply. If the upper bound of a measure is not defined, e.g. “blood glucose measurement daily for at least 2 weeks”, the upper bound will be captured as “INF” (infinite). Infinite as an upper bound is never included. If "true" the upper bound is part of the interval. **Invariant**: if (*upperBound* == INF) { *includeUpperBound* = false }.
semantic LogicalExpression	[1...1]	Measure semantic represents a unit of measure or scale specified by the interval values. It is described using a logical expression using standard-based terminology (i.e. SNOMED CT). For systolic blood pressure, the unit of measure is millimeters of mercury, and thus the measure semantic is a SNOMED CT concept: 259018001 |Millimeter of mercury (qualifier value). For blood glucose measurement daily for 2 weeks, the measure semantic would be “258705008 |week (qualifier value)”. The semantic is also used to specify that the result is a countable quantity of findings or phenomena described in the *topic*.

Attribute	Multiplicity	Notes
		• For example, if the **topic** is "retinal hemorrhages" with a **result** of three (i.e. [3,3]) the **result** semantic of "countable quantity" specifies that *exactly three hemorrhages were counted/observed*. If Measure is used to represent date or time: • Date/time using Unix Epoch time: [762636008] Duration, [257997001] Seconds • Duration using Unix Epoch time start time and end time: [762636008] Duration, [257997001] Seconds

2.1.3.2. Repetition

This class builds on Measure and it is used to represent when an action is requested for more than a single occurrence. Repetition is an optional component for a RequestCircumstance.

Attribute	Multiplicity	Notes
periodStart Measure	[1...1]	This required field is used to represent when a repeated action should begin (e.g. NOW). If it is not specified, a default value of [0,INF) will be used. • NOW is represented as a interval of time values. If the timing is precise the lower and upper bounds may be identical; otherwise the interval would match the precision of the original time observations.
periodDuration Measure	[1...1]	This required field is used to represent how long a repeated action should persist (e.g. for a year). If it is not specified, a default value of [0,INF) will be used.
eventSeparation Measure	[1...1]	This required field is used to represent how long between actions (e.g. 1 week). If it is not specified, a default value of [0,INF) will be used.
eventDuration Measure	[0...1]	This optional field is used to represent how long a repetition should persist (e.g. for 2 hours). If it is not specified, a default value of [0,INF) will be used.
eventFrequency Measure	[1...1]	This required field is used to represent how often the action should occur (e.g. 4 times per month). If it is not specified, a default value of [0, INF) will be used.

2.1.3.3. LogicalExpression

This class represents a wrapper for logical expression.

Attribute	Multiplicity	Notes
expression Expression	[1...1]	Logical expression could be represented using FHIR Expression structure or a similar standard-based syntax (e.g. *SNOMED CT Expression Constrain Language - ECL*). The expression must use valid, standard-based terminology.

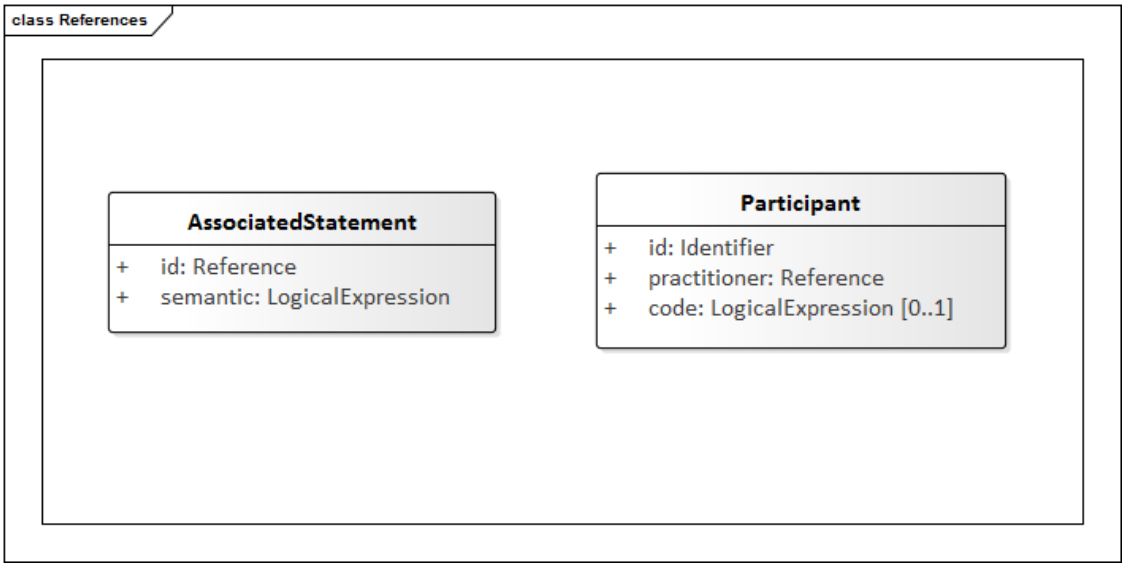
2.1.4. References

A clinical statement references other information managed by a system:

- references to patient/records
- references to health practitioners

ANF statements may also reference other related statements

Figure 2.4. References



2.1.4.1. AssociatedStatement

This class specifies how a statement may be associated with another statement.

Note

Associated statements can be used to represent a complex model consisting of a set of associated statements (see ???).

Attribute	Multiplicity	Notes
id Reference	[1...1]	A reference to the associated statement.
semantic LogicalExpression	[1...1]	A logical expression to capture how the target statement is associated (e.g. a precondition, an interpretation, a component).

2.1.4.2. Participant

This class specifies the role/specialties/services that a practitioner may perform relative to the ANFStatement:

- the author
- requested participant
- performance participant

Attribute	Multiplicity	Notes
id Identifier	[1...1]	Unique identifier (e.g. National Provider Identifier).
practitioner Reference	[1...1]	Reference to the participating practitioner.
code LogicalExpression	[0...1]	Role(s) which this practitioner is authorized to perform for the organization.

2.2. Editorial Rules

The editorial rules outlined below provide criteria for disambiguating and removing redundancy between topic versus result, performance versus request:

2.2.1. General Editorial Rules

The most important editorial rule for ANF statements is to first decide whether something is being requested or performed. In addition to this there are other general editorial rules that apply to all ANF statements regarding timing, subject of information and the ability to associate related statements.

Editorial Rule 2.1. Performance versus request

- This rule mandates that an ANF Statement must describe either the *performance of an action* or the *request for an action*.
  - A Performance may include the passive observation of a phenomenon related to patients and their health status or family history, and may also include active interventions, such as providing education or administering medications or documenting that a patient is participating in exercise to improve their overall health status.
  - A Request may include requests for clinical testing, active interventions, future goals, or consultation with other providers.

Editorial Rule 2.2. Timing - past, present, or future

- For a Performance of Action, the Timing can represent a time in the past or a current time. If a history of a performance of action is to be represented in ANF the Timing will be for a time in the past prior to the statement. Otherwise the Timing will be represented with the current time of the statement.
- For a Request of Action, the Timing will always represent a future time.

Editorial Rule 2.3. Related statements should be associated

- Use an associated statement when it is important for the interpretation of one statement that the other statements were observed, performed, or requested. Also, if there is some implicitness that the two statements are related (pleural empyema with fistula) or that they are unrelated (Akinetic seizure without atonia) then the two statements should be associated.

Editorial Rule 2.4. Subject of information is used to represent who the statement is about

- The subjectOfInformation is used to represent who the statement is about. This is normally the patient (Subject of record) unless explicitly stated otherwise, for example Mother, Sibling, Donor, etc.

2.2.2. Topic Editorial Rules

Editorial Rule 2.5. Topics are always an action

- The particulars of how topics—and other statement fields—are modeled as a Terminology Layer concern, not a Statement Layer concern. The Statement Layer does require that the Terminology Expression fields in a statement are disjoint: There should be no confusion—or creation of false dichotomies. There should be one, and only one, place to put each type of information in a terminology expression. For example, the Statement Layer defines a particular place to represent the subject of information. Therefore, the Terminology Layer must not allow the subject of information to be redundantly—and possibly contradictory—represented in a topic expression (such as would be the case if "maternal history of diabetes" were an allowed topic expression). The Statement Layer requires that the topic represent an Action as a code or expression according to the rules of the Terminology Layer, and that the rules of the Terminology Layer enforce a disjointness between different types of terminology expressions. Here we present a starting point for what the Terminology Layer editorial rules may look like, based on current SNOMED CT practice.

- SNOMED CT can accommodate this requirement for simple observations by using Observation procedure to represent the topic (or other types of procedures when appropriate, such as the administration of a medication). In SNOMED CT examples, the Observation procedure specifies a Has focus attribute linking it to the Clinical Finding or Disorder that it is being observed. The observation procedure can also be further refined by adding attributes in the terminology model, including Method, Procedure site - Direct, (if appropriate) Laterality, and Using device.

- Medication administrations will use an Administration of substance concept to represent the topic. All Administration of substance concepts will be refined with the substance, dose form and strength being requested. If Route of administration exists, then it will also be added.

- Laboratory tests will use a Laboratory Procedure concept to represent the topic. These concepts can be further refined.

- Imaging Procedures will use an Imaging Procedure concept to represent the topic. These concepts will be further refined with a Method, Procedure site and (if appropriate) a laterality for those sites that are lateralizable.

Editorial Rule 2.6. Prerequisites must be separated from the topic

- A prerequisite must be separable from the topic and should be expressed as a stand-alone clinical statement.
- A prerequisite is a state that must exist before something else can happen or be done. Prerequisites are part of the details under which a procedure is being performed. The state must exist prior to the performance of the action.

Editorial Rule 2.7. Separate compound topics

- For the purposes of ANF, a statement is a request or performance of an action that should exist independently. Thus, if a compound topic contains two topics that could each exist separately, then they should be divided into separate ANF Statements. These independent ANF Statements can then be associated with each other as associated statements.
- For example, "Negative screen for PTSD and depression", contains two separate ANF Statements that would then be associated to each other. However, if the narrative represents two or more actions that are performed as a single activity at the same time without the need for stopping the action, then a single topic would be used. For example, "Lumbar/Thoracic Spine CT" would be represented with a single topic as it represents a single activity that is performed at the same time even though a Lumbar CT and a Thoracic CT could be done separately.

Editorial Rule 2.8. Techniques are inseparable from the topic

- A technique must be true within the duration of the performance.
- A technique is inseparable from the topic and cannot be expressed as a stand-alone clinical statement.
- A technique is a device used, a method applied, or a temporary state in which the patient was actively placed during performance of the action.

2.2.3. Circumstance Editorial Rules

Editorial Rule 2.9. Results are always a ranged quantity

- Results are always a Measure, which is a ranged quantity. Measure includes both a numeric interval along with a Measure Semantic specified as a Logical Expression.
- If a Result is intended to represent a numeric result then the upperBound and lowerBound would be populated with the appropriate numeric values and the Measure Semantic would indicate the unit of measure.

Editorial Rule 2.10. Presence and absence are a countable quantity

- Any statement that represents the Presence or implies Presence of a Topic will have a Result with an upperBound of infinite (inf), lowerBound of 1, and result semantic of "Countable quantity".
- Any statement that represents the Absence or implies Absence of a Topic will have a Result with an upperBound of 0, lowerBound of 0, and result semantic of "Countable quantity".

Editorial Rule 2.11. Participants can be specified or requested

- A Performance of action can specify participants using participant in PerformanceCircumstance.
- A Request for action can specify requested participants using requestedParticipant in RequestCircumstance.

Editorial Rule 2.12. Purpose indicates the reason for a request or performance

- The purpose is why an action was requested or performed. The purpose of the topic is typically some type of therapeutic intent, diagnostic intent, or both. There can be more than one therapeutic intent and diagnostic intent. While the purpose can also exist as a separate clinical statement, if you specifically want to state that a action was performed for a particular purpose, it must be represented using the purpose.

2.2.4. Performance Circumstance Editorial Rules

Editorial Rule 2.13. Status indicates the state of a result

- The status of a Performance of action can be specified with concepts such as "on hold", "completed", "rejected", etc.

Editorial Rule 2.14. healthRisk indicates the clinical risk of the result

- In PerformanceCircumstance, healthRisk is used to flag a result with coded values such as 'low', 'normal', high', and 'critical'.

Editorial Rule 2.15. reference can be specified for a result

- In Performance Circumstance "referenceRange" is the interval of values that are normal for the observation/finding described by the "topic" for this "subject". It refers to "normal" for the patient/subject under specific conditions.

2.2.5. Request Circumstance Editorial Rules

Editorial Rule 2.16. Priority defaults to routine for a request

- Priority is used to represent the priority for which a request is to be carried out. By default a Request will be considered "routine" unless otherwise specified.

Editorial Rule 2.17. Repetition is used to request multiple occurrences of the thing described in the topic

- Repetition is used to represent when an action is requested for more than a single occurrence.

Editorial Rule 2.18. A desired result can be specified in a request

- A desired result can be specified as a Measure using requestedResult in RequestCircumstance.
- If a requestedResult is specified, the appropriate upperBound and lowerBound is specified with the correct result semantic.
- If a requestedResult is unspecifiedt, the value is set to [0, inf) with a result semantic of "Countable quantity".

3. How ANF Works: ANF Clinical Statements

In the context of the ANF Model, a *clinical statement* represents an entry in the patient record that documents, in a structured/computable manner, clinical information related to the patient that is asserted by a particular source, recorded, and potentially verified.

Clinical Input Forms allow EHR systems to capture and store clinical statements in multiple ways based on use cases, local preferences, etc. ANF strives to standardize the structure of clinical statements to eliminate the disparity of clinical information by introducing a common, normalized form. ANF can be used as consistent transformation target for the multiple differing clinical information representations that currently exist, making this clinical information more easily computable. Thus it eliminates the need to create multiple ways to analyze the same data.

3.1. Types of *ANF Statements*

ANF statement may express a **performance** or a **request**; the criteria for differentiating them for are described as **editorial rules**: Performance versus request

- This rule mandates that an ANF Statement must describe either the *performance of an action* or the *request for an action*.
    - A Performance may include the passive observation of a phenomenon related to patients and their health status or family history, and may also include active interventions, such as providing education or administering medications or documenting that a patient is participating in exercise to improve their overall health status.
    - A Request may include requests for clinical testing, active interventions, future goals, or consultation with other providers.

3.1.1. Performance of Action Statements

A Performance of Action statements describe an action performed, and any result associated with the action specified by the topic.

The following are examples of performance statements:

- The presence or absence of a clinical finding
    - Diabetes mellitus is present
    - Diabetes mellitus is not present
    - Retinal hemorrhage is present
- The results of a specific test/screening or procedure
    - Pulse Rate 68 bpm, taken by pulse oximeter
    - Systolic blood pressure 120 mmHg, taken on right brachial artery, using BP cuff adult size, patient in sitting position for at least 5 minutes, urinated not more than 30 minutes prior to measurement
    - Three retinal hemorrhages
    - Positive screen for fall risk
    - Negative screen for PTSD and depression
- Administered a medication or other substance
    - Patient took one Acetaminophen 500 mg tablet by mouth for pain

- Provision of educational materials
  - Patient was provided with educational materials on diabetes

- he presence of any other state or specific characteristic that is clinically relevant

  - Family history of breast cancer

3.1.1.1. Presence or Absence of a Clinical Phenomenon

To describe the presence or absence of a phenomenon, an ANF statement requires a topic, timing, and a result defined based on editorial rules:

Topics are always an action

- The particulars of how topics—and other statement fields—are modeled as a Terminology Layer concern, not a Statement Layer concern. The Statement Layer does require that the Terminology Expression fields in a statement are disjoint: There should be no confusion—or creation of false dichotomies. There should be one, and only one, place to put each type of information in a terminology expression. For example, the Statement Layer defines a particular place to represent the subject of information. Therefore, the Terminology Layer must not allow the subject of information to be redundantly—and possibly contradictory—represented in a topic expression (such as would be the case if "maternal history of diabetes" were an allowed topic expression). The Statement Layer requires that the topic represent an Action as a code or expression according to the rules of the Terminology Layer, and that the rules of the Terminology Layer enforce a disjointness between different types of terminology expressions. Here we present a starting point for what the Terminology Layer editorial rules may look like, based on current SNOMED CT practice.

- SNOMED CT can accommodate this requirement for simple observations by using Observation procedure to represent the topic (or other types of procedures when appropriate, such as the administration of a medication). In SNOMED CT examples, the Observation procedure specifies a Has focus attribute linking it to the Clinical Finding or Disorder that it is being observed. The observation procedure can also be further refined by adding attributes in the terminology model, including Method, Procedure site - Direct, (if appropriate) Laterality, and Using device.

- Medication administrations will use an Administration of substance concept to represent the topic. All Administration of substance concepts will be refined with the substance, dose form and strength being requested. If Route of administration exists, then it will also be added.

- Laboratory tests will use a Laboratory Procedure concept to represent the topic. These concepts can be further refined.

- Imaging Procedures will use an Imaging Procedure concept to represent the topic. These concepts will be further refined with a Method, Procedure site and (if appropriate) a laterality for those sites that are lateralizable.

Timing - past, present, or future

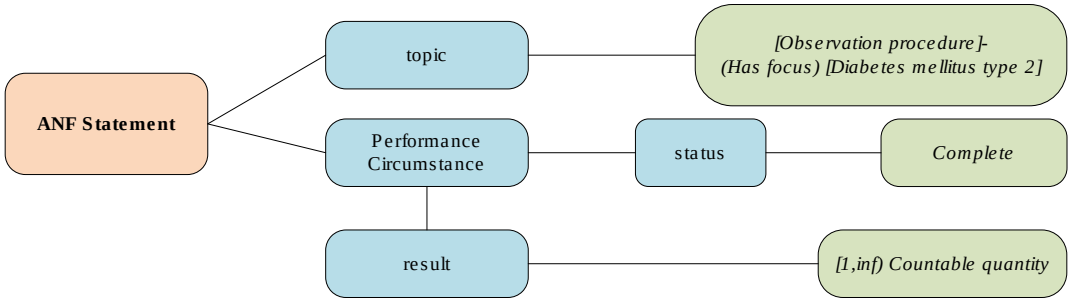
- For a Performance of Action, the Timing can represent a time in the past or a current time. If a history of a performance of action is to be represented in ANF the Timing will be for a time in the past prior to the statement. Otherwise the Timing will be represented with the current time of the statement.

- For a Request of Action, the Timing will always represent a future time.

- Any statement that represents the Presence or implies Presence of a Topic will have a Result with an upperBound of infinite (inf), lowerBound of 1, and result semantic of "Countable quantity".

- Any statement that represents the Absence or implies Absence of a Topic will have a Result with an upperBound of 0, lowerBound of 0, and result semantic of "Countable quantity".

Figure 3.1. Diabetes Mellitus Present ANF Example

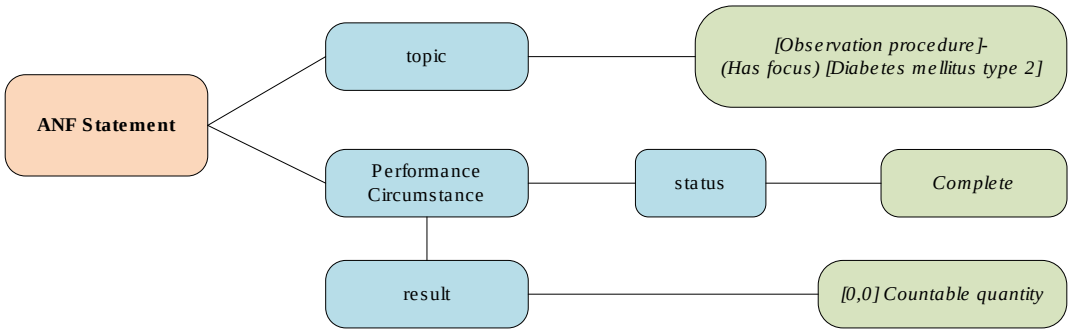


Diabetes Mellitus Type 2 Present.

In the Diabetes Mellitus type 2 example above, the Topic is an Observation procedure with a Has focus of Diabetes mellitus type 2. To represent that it is present, the Result is a lowerBound of 1, an upperBound of infinite (inf), and a measure semantic of "Countable quantity". For more details on the syntax used to represent a Result see the Measure Class here: Section 2.1.3.1, “Measure ”To see a more detailed representation of the Diabetes Mellitus Type 2 Present example see the tabular form here: ???

See Editorial Rule: Presence and absence are a countable quantity

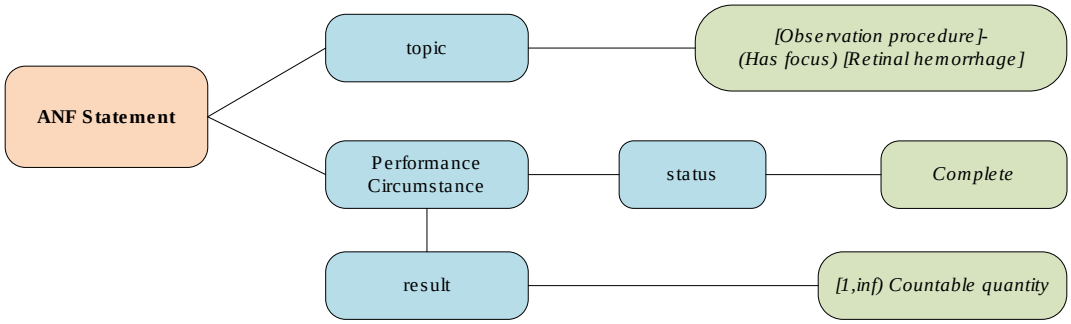
Figure 3.2. Diabetes Mellitus Type 2 Absent ANF Example



Diabetes Mellitus Type 2 Absent.

In the Diabetes Mellitus Type 2 Absent example, the topic is the same as Diabetes Mellitus Type 2 Present example. The difference is in the Result which is represented as an upperBound and lowerBound of zero with the same measure semantic. To see a more detailed representation see the tabular form here: ???

Figure 3.3. Retinal Hemorrhage Present ANF Example



Retinal Hemorrhage Present.

To see a more detailed representation see the tabular form here: ???

3.1.1.2. Test/Screening or Procedure and Resultant Value

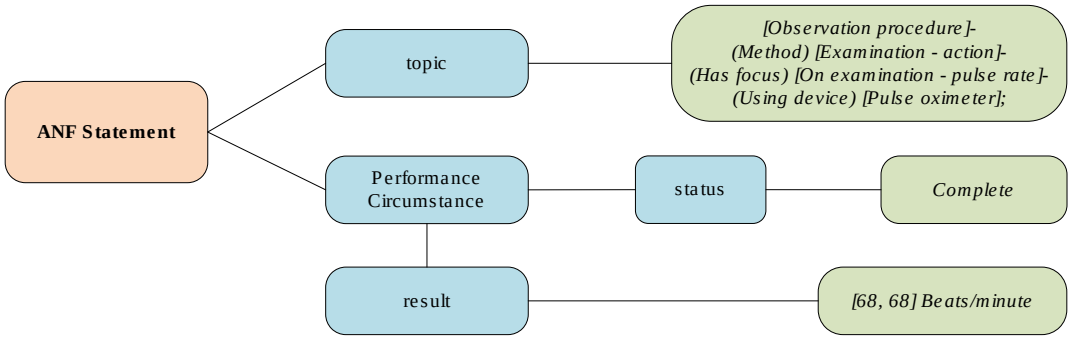
See Editorial Rule: Results are always a ranged quantity

- Results are always a Measure, which is a ranged quantity. Measure includes both a numeric interval along with a Measure Semantic specified as a Logical Expression.
- If a Result is intended to represent a numeric result then the upperBound and lowerBound would be populated with the appropriate numeric values and the Measure Semantic would indicate the unit of measure.

See Editorial Rule: Techniques are inseparable from the topic

- A technique must be true within the duration of the performance.
- A technique is inseparable from the topic and cannot be expressed as a stand-alone clinical statement.
- A technique is a device used, a method applied, or a temporary state in which the patient was actively placed during performance of the action.

Figure 3.4. Pulse Rate ANF Example



Pulse Rate 68 bpm, Taken by Pulse Oximeter.

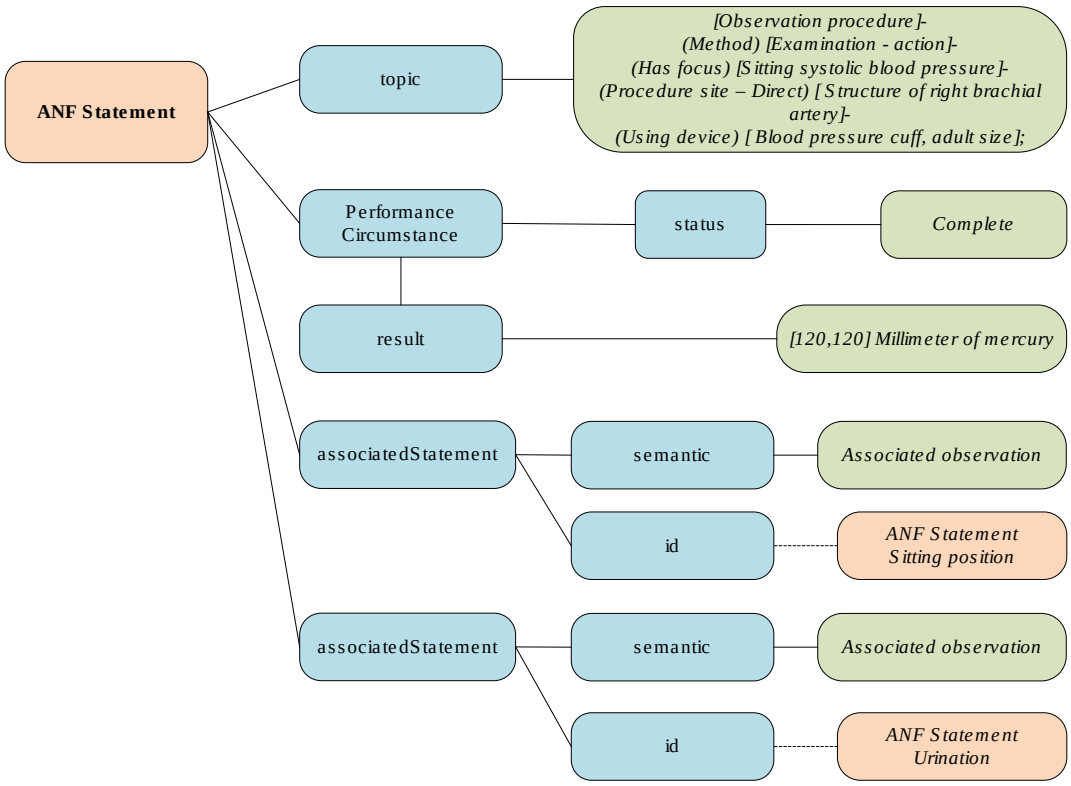
The Pulse Rate example above utilizes a technique, the pulse oximeter device, and contains a resultant value of 68 beats/minute. Since a Result is represented with an upperBound and lowerBound they are both represented as 68 in this case. To see a more detailed representation see the tabular form here: ???

See Editorial Rule: Prerequisites must be separated from the topic

- A prerequisite must be separable from the topic and should be expressed as a stand-alone clinical statement.

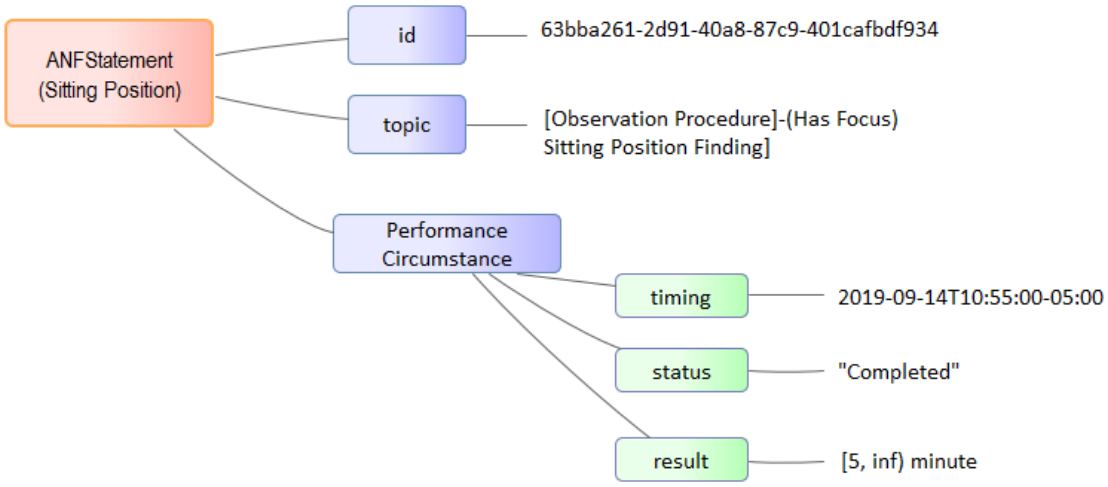
- A prerequisite is a state that must exist before something else can happen or be done. Prerequisites are part of the details under which a procedure is being performed. The state must exist prior to the performance of the action.

Figure 3.5. Systolic Blood Pressure with Associated Statements ANF Example



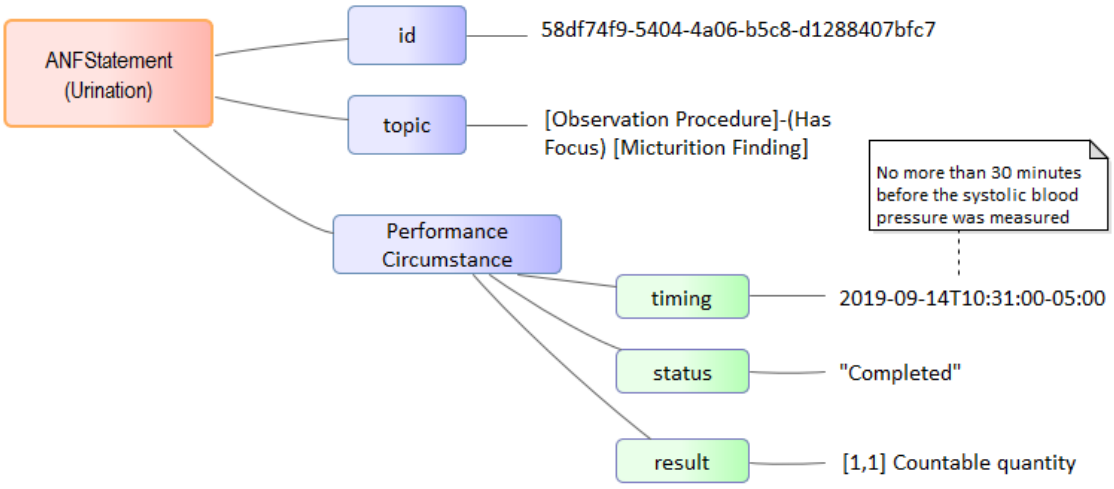
Systolic Blood Pressure 120 mmHg, Taken on Right Brachial Artery, Using BP Cuff Adult Size, Patient in Sitting Position for at Least 5 Minutes, Urinated Not More Than 30 Minutes Prior to Measurement.

Figure 3.6. Systolic Blood Pressure Sitting Position Associated ANF Statement Example



Patient in Sitting Position for at Least 5 Minutes.

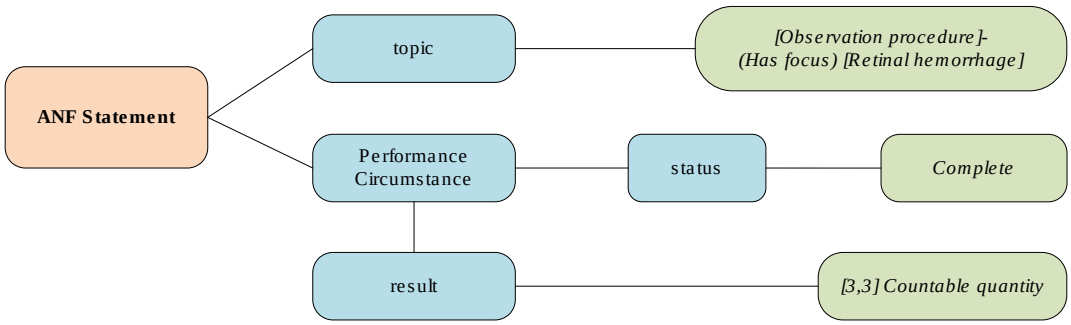
Figure 3.7. Systolic Blood Pressure Urination Associated ANF Statement Example



Urinated Not More Than 30 Minutes Prior to Measurement.

The systolic blood pressure example above not only includes a technique of using an adult sized cuff, but also includes two prerequisites that are represented as separate associated ANF Statements. In the Associated Statements we see examples of Results having a range of values using the upperBound and lowerBound. Additionally, while the narrative does not explicitly state that the blood pressure is taken in the sitting position, it is implied by the prerequisite that it is taken in the sitting position based on the prerequisite that the patient was in the sitting position for at least 5 minutes. To see a more detailed representation see the tabular form here: ???

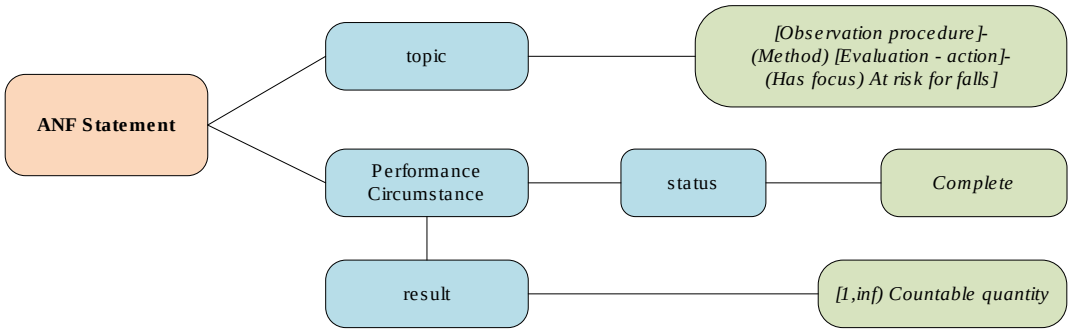
Figure 3.8. Three Retinal Hemorrhages ANF Example



Three Retinal Hemorrhages.

To see a more detailed representation see the tabular form here: ???

Figure 3.9. Positive Screen for Fall Risk ANF Example



Positive Screen for Fall Risk.

See Editorial Rule: Separate compound topics

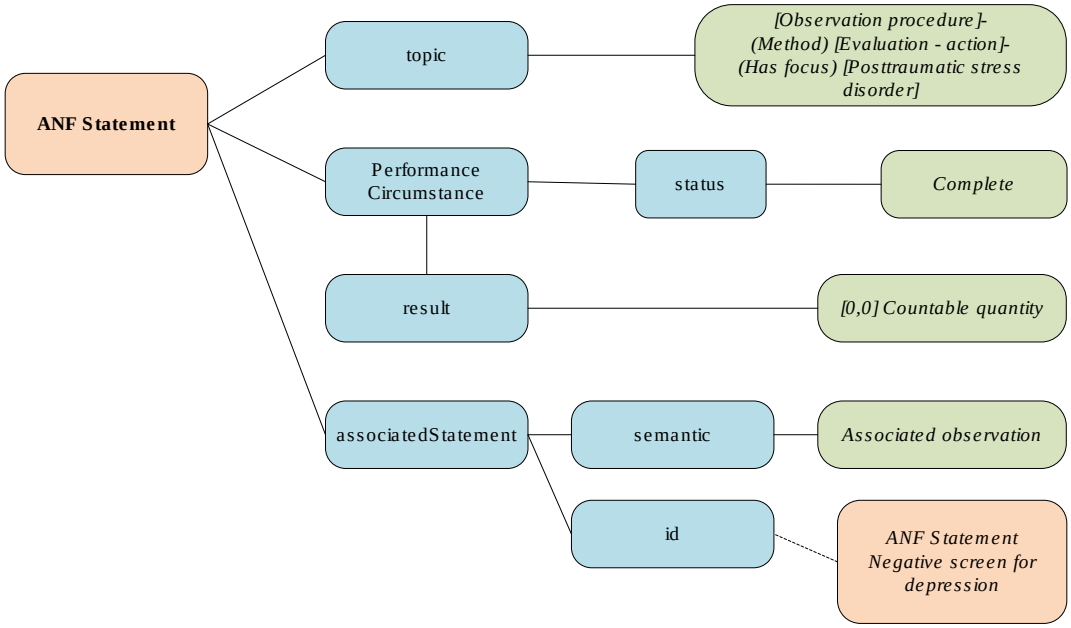
- For the purposes of ANF, a statement is a request or performance of an action that should exist independently. Thus, if a compound topic contains two topics that could each exist separately, then they should be divided into separate ANF Statements. These independent ANF Statements can then be associated with each other as associated statements.

- For example, "Negative screen for PTSD and depression", contains two separate ANF Statements that would then be associated to each other. However, if the narrative represents two or more actions that are performed as a single activity at the same time without the need for stopping the action, then a single topic would be used. For example, "Lumbar/Thoracic Spine CT" would be represented with a single topic as it represents a single activity that is performed at the same time even though a Lumbar CT and a Thoracic CT could be done separately.

See Editorial Rule: Related statements should be associated

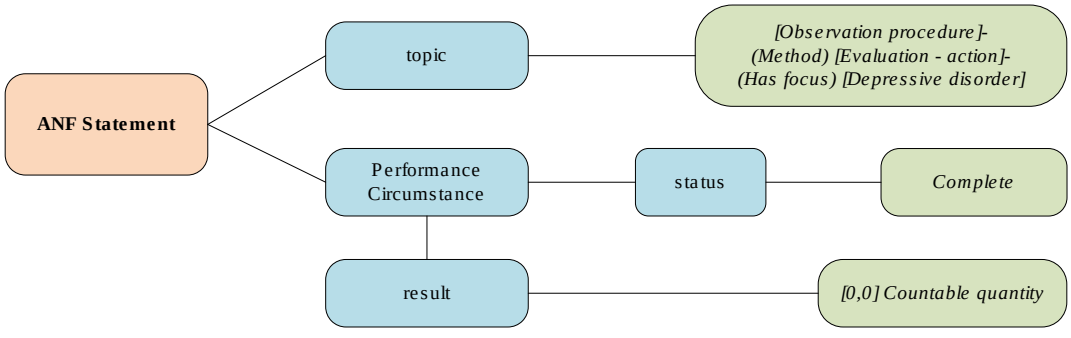
- Use an associated statement when it is important for the interpretation of one statement that the other statements were observed, performed, or requested. Also, if there is some implicitness that the two statements are related (pleural empyema with fistula) or that they are unrelated (Akinetic seizure without atonia) then the two statements should be associated.

Figure 3.10. Negative Screen for PTSD ANF Example



Negative Screen for PTSD.

Figure 3.11. Negative Screen for Depression ANF Example



Negative Screen for Depression.

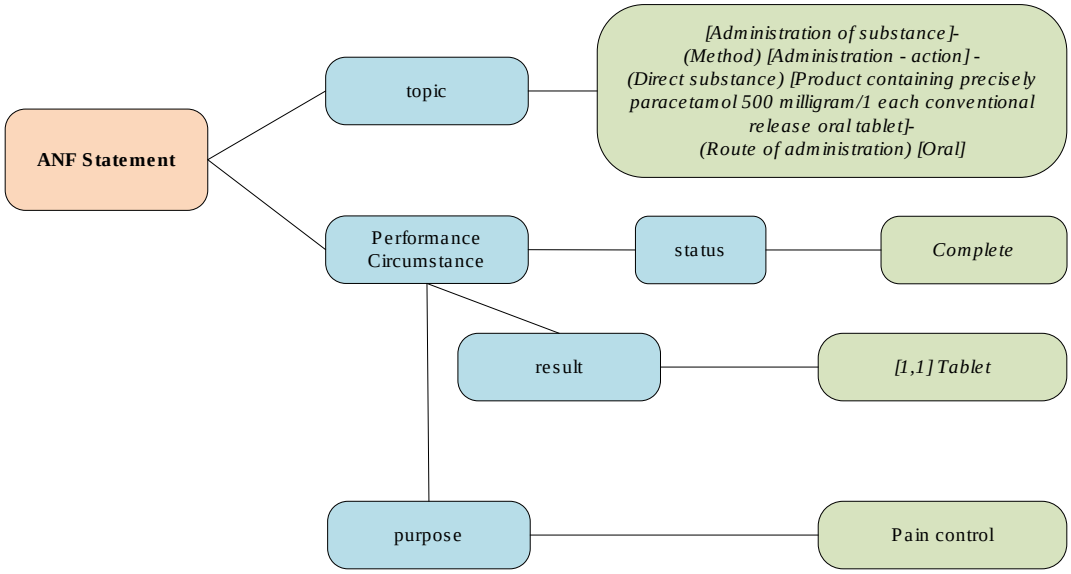
3.1.1.3. Administering a Medication or Other Substance

See Editorial Rule: Topics are always an action

See Editorial Rule: Purpose indicates the reason for a request or performance

- The purpose is why an action was requested or performed. The purpose of the topic is typically some type of therapeutic intent, diagnostic intent, or both. There can be more than one therapeutic intent and diagnostic intent. While the purpose can also exist as a separate clinical statement, if you specifically want to state that a action was performed for a particular purpose, it must be represented using the purpose.

Figure 3.12. Administration of Medication ANF Example



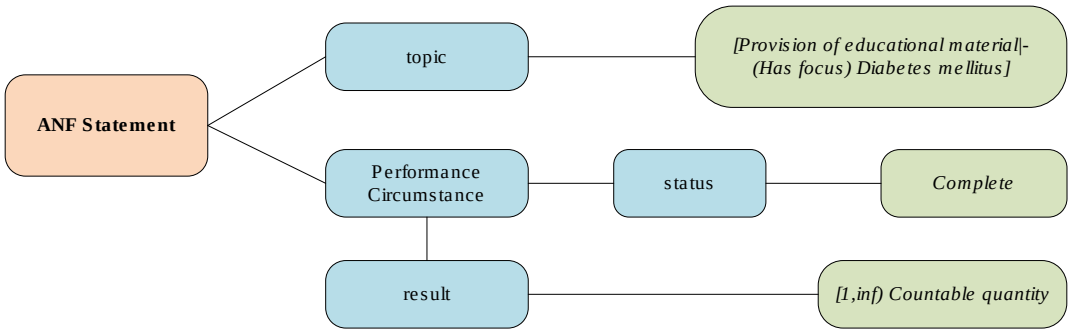
Patient Took One Acetaminophen 500 mg Tablet by Mouth for Pain.

In the medication example above a purpose is specified using Pain control which has the **purpose** of **pain control**. The Topic is built using Administration of substance with a Direct substance specifying the pharmaceutical product including the strength and dose form and a Route of Administration specifying Oral. In addition to the dose form in the Topic, the dose form is also specified in the Measure semantic for the Result. This allows for the specification of multiple or partial dose forms.

3.1.1.4. Provision of Educational Materials

See Editorial Rule: Topics are always an action

Figure 3.13. Provision of Educational Material ANF Example

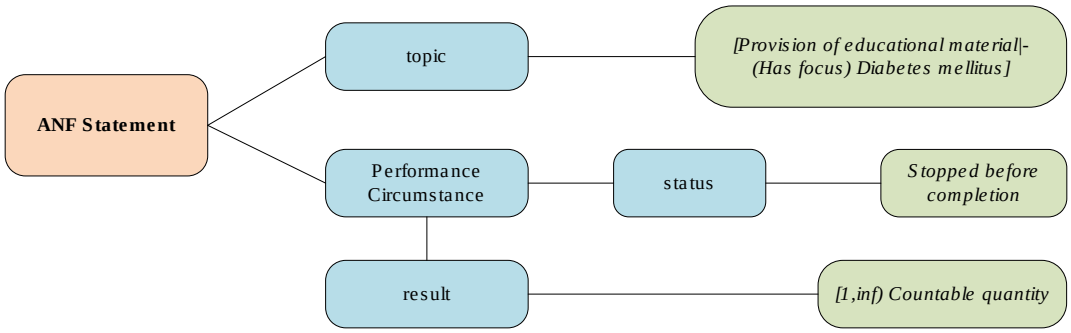


Patient was Provided with Educational Material on Diabetes.

In this example, the concept Provision of educational material is used with a Has focus of Diabetes mellitus.

See Editorial Rule: Status indicates the state of a result

Figure 3.14. Provision of Educational Material Stopped Before Completion ANF Example



Patient was Provided with Educational Material on Diabetes but Education was Stopped Before Completion.

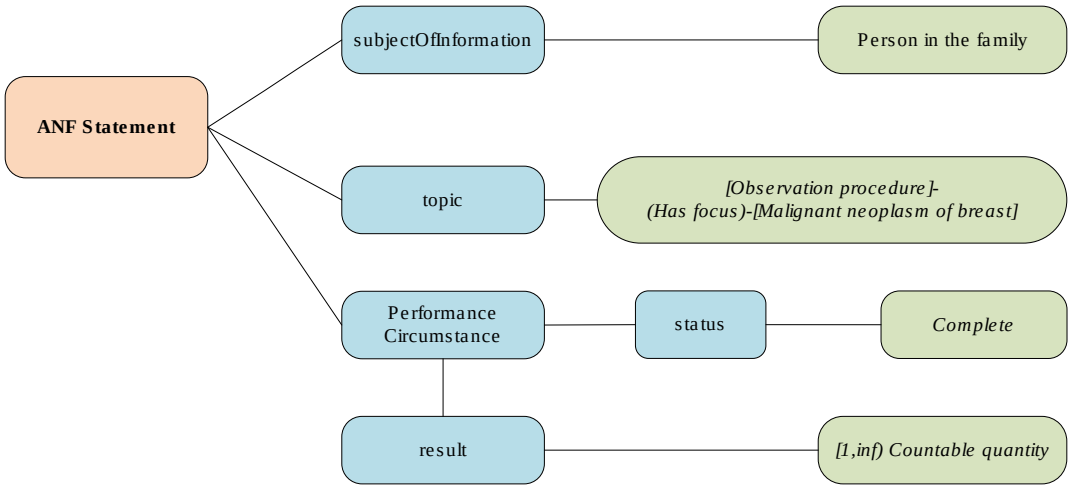
This example is similar to the prior example of completing the education, however the status is used to represent the education was not completed after starting.

3.1.1.5. Other States or Specific Characteristics That Are Clinically Relevant

See Editorial Rule: Subject of information is used to represent family and donor history

- The subjectOfInformation is used to represent who the statement is about. This is normally the patient (Subject of record) unless explicitly stated otherwise, for example Mother, Sibling, Donor, etc.

Figure 3.15. Family History ANF Example



Family History of Breast Cancer.

In the Family history of breast cancer example we see that the Family history is represented by the Subject of information with a value of Person in the family.

See Editorial Rule: Reference Range can be specified for a result

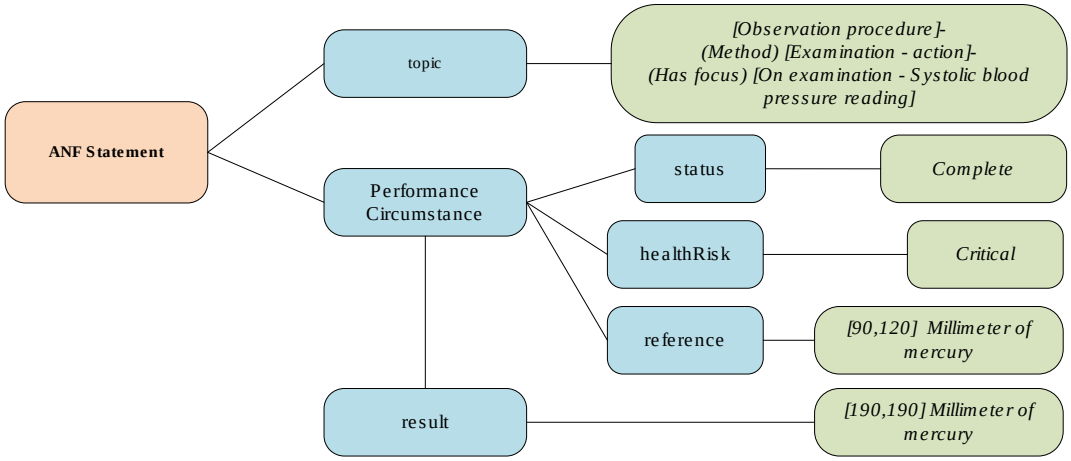
- In Performance Circumstance "referenceRange" is the interval of values that are normal for the observation/finding described by the "topic" for this "subject". It refers to "normal" for the patient/subject under specific conditions.

See Editorial Rule: HealthRisk indicates the clinical risk of the result

- In PerformanceCircumstance, healthRisk is used to flag a result with coded values such as 'low', 'normal', high', and 'critical'.

3.1.1.6. Reference Range Information or Health Risk Specified

Figure 3.16. Systolic Blood Pressure with Reference Range and Health Risk ANF Example



Systolic Blood Pressure 190 mmHg, Reference Range (90-120), Health Risk Critical.

Systolic Blood Pressure for adults has a reference range of 90-120 and is represented in the reference. The reference utilizes the same Measure class that the Result utilizes and the syntax represented in the image above is described in detail here: Section 2.1.3.1, “Measure ”. Systolic Blood pressure above 180 would represent a critical health risk and is represented in the healthRisk.

3.1.2. Request Clinical Statements

A Request for Action clinical statement describes a request made by a clinician. Most of the times, but not always, the object of the request (e.g., laboratory test, medication order) will be fulfilled by someone other than the clinician (e.g., laboratory professional, pharmacist) making the request. All information about the request will be documented in this clinical statement, including information about details relating to the request, such as patient must fast for 12 hours before having a lipids blood test.

Examples of Request clinical statements:

- Request for Rheumatoid factor 1 time routine
- Request for X-ray chest to evaluate for heart failure
- Cardiology referral
- Ribavirin 200 mg capsule oral, take 2 capsules every morning
- Advised to participate in tobacco cessation counseling once a week.

3.1.2.1. Request Examples

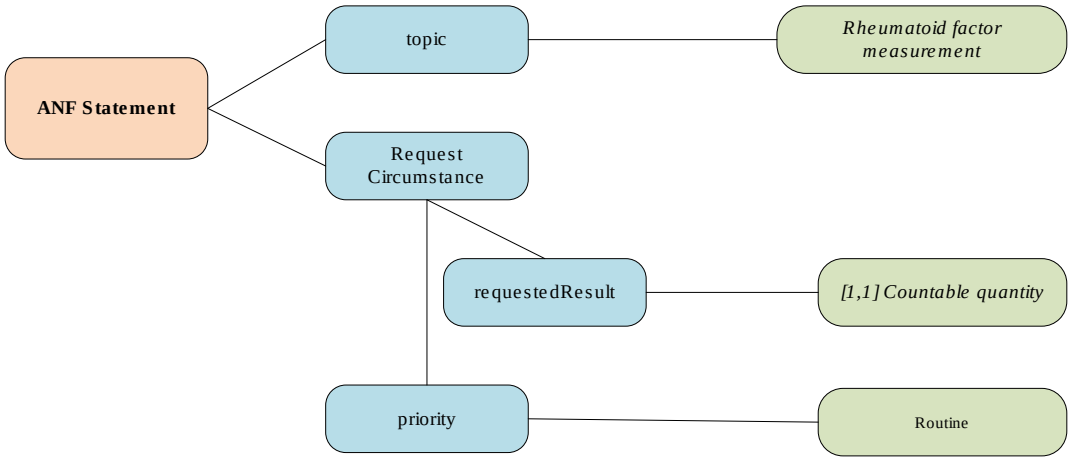
See Editorial Rule: Timing - past, present, or future

See Editorial Rule: Topics are always an action

See Editorial Rule: Priority defaults to routine for a request

- Priority is used to represent the priority for which a request is to be carried out. By default a Request will be considered "routine" unless otherwise specified.

Figure 3.17. Laboratory Request ANF Example

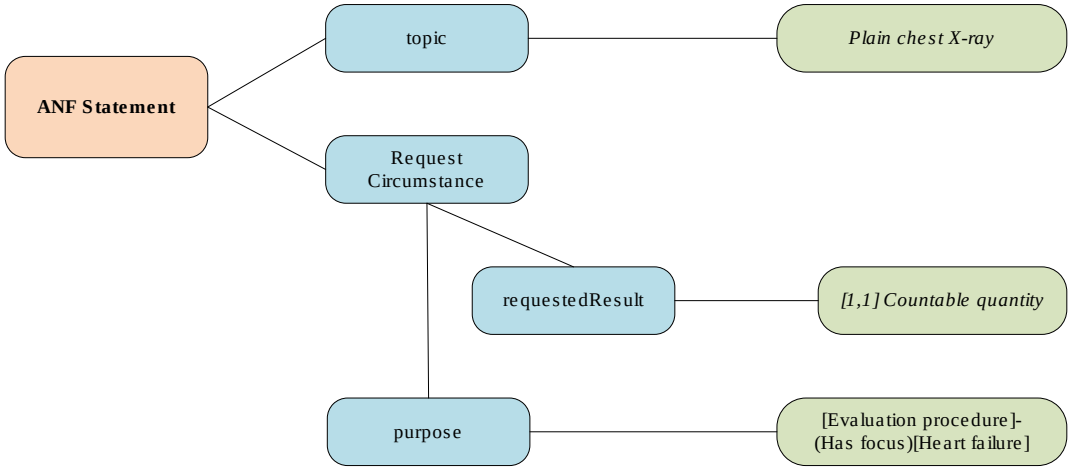


Rheumatoid Factor 1 Time Routine.

The Laboratory Request example above shows how the topic is built using a laboratory procedure concept, with no refinements in this case. It also has a Priority of Routine as stated in the narrative description. The requestedResult in this example is used to represent that you are requesting a single measurement be performed.

See Editorial Rule: Topics are always an action

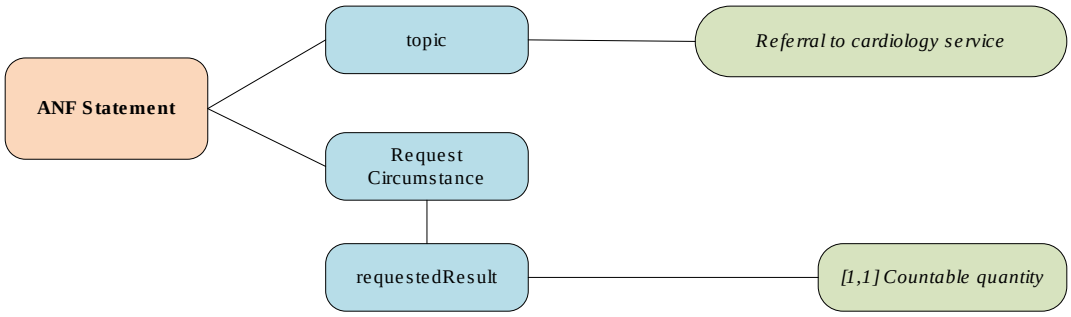
Figure 3.18. Imaging Request ANF Example



X-ray Chest to Evaluate for Heart Failure.

The Imaging Request example above is built using a subtype of image procedure concept and includes a Purpose to record why the procedure is being done.

Figure 3.19. Referral Request ANF Example



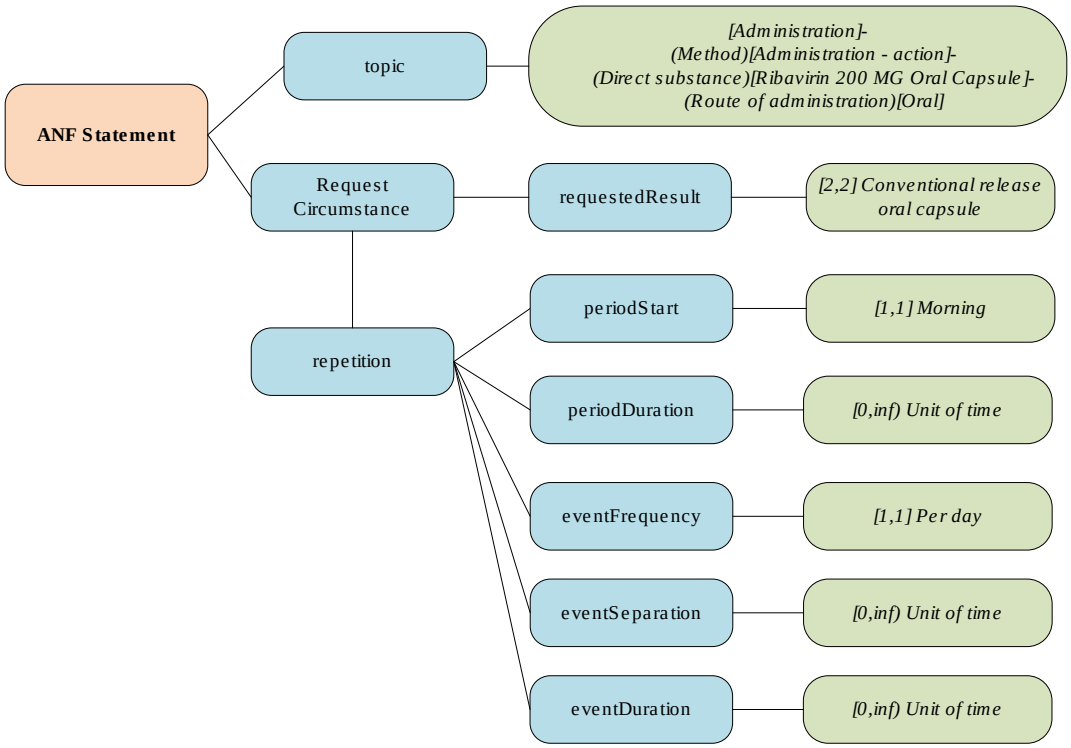
Cardiology Referral.

See Editorial Rule: Topics are always an action

See Editorial Rule: Repetition is used to request multiple occurences of a topic

- Repetition is used to represent when an action is requested for more than a single occurrence.

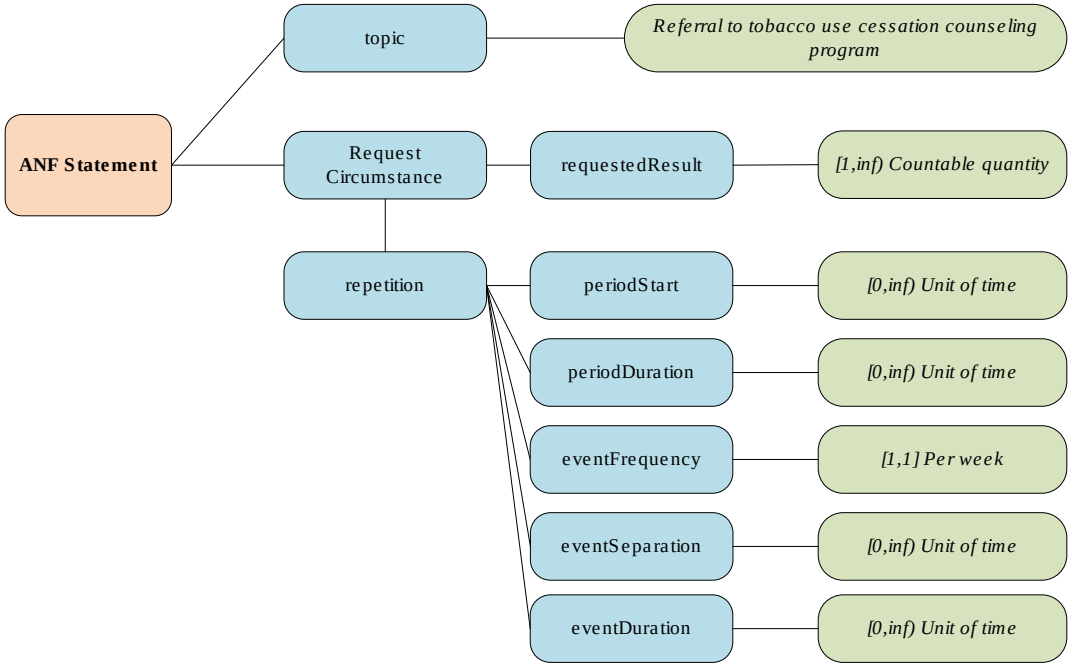
Figure 3.20. Medication Request ANF Example



Ribavirin 200 mg Capsule Oral, Take 2 Capsules Every Morning.

The Medication request example represents one of the more complicated ANF Statements that includes not only the Topic, but also the Repetition information for completing the request. The Topic is built using Administration of substance with a Direct substance specifying the pharmaceutical product including the strength and dose form and a Route of Administration specifying Oral. In addition to the dose form in the Topic, the dose form is also specified in the Measure semantic for the requestedResult. This allows for the specification of multiple or partial dose forms.

Figure 3.21. Counseling Request ANF Example



Advised to Participate in Tobacco Cessation Counseling Once a Week.

In this example we see Repetition used only to define the eventFrequency while the other Repetition information is defaulted to [0,inf) Unit of time.

4. Methodology—ANF Design Principles and Rules

4.1. ANF Design Principles

As an overarching principle we favor the simpler, consistent model over more complex models that allow for multiple inconsistent representations. As such, the following principles have been used when designing the ANF model:

A. **Overall Model Simplicity:** In cases where different principles collide, we shall favor simplicity of the entire system over simplicity in one area of the system. This principle is achieved by avoiding complex types, inheritance/derivations, and extensions. Instead, ANF relies on sophisticated terminology and a polymorphic information model (see ANF Reference Model) that can be used to create complex models of interrelated statements (see Wound Assessment) and can use platform-specific primitive types. Simplicity is the direct result of normalization and maintaining a minimum set of data required to express clinical statements.

B. **Convention Over Configuration:** Convention over configuration is a design paradigm used by frameworks that decreases the number of decisions that a developer using the framework is required to make, without necessarily losing flexibility because conventions can be overridden when necessary.

C. **Model Consistency:** Patterns should allow the consistent representation of information that is commonly shared across models. For instance, attribution and participant information should be captured consistently. Failure to do so forces implementers to develop heuristics to capture and normalize attribution information that is represented or extended differently in different classes (see ??? and ???).

D. **No Semantic Overloading:** Semantic overloading occurs when a model attribute's meaning changes entirely, depending on context. While the refinement of the semantics of an attribute in a subclass is acceptable, a change of meaning is problematic. For instance, in FHIR, the Composition class defines an attribute called Subject. In some subclasses, the attribute may be the entity that this composition refers to (e.g., the patient in a medical record). In other cases, it is the topic being discussed by the composition (e.g., a medication orderable catalog).

E. **Assumption-free:** Implied semantics must be surfaced explicitly in the model.

F. **Composition Over Inheritance:** Composition over inheritance (or composite reuse principle) is the principle that classes should achieve polymorphic behavior and code reuse by their composition (by containing those instances of other classes that implement the desired functionality) rather than inheritance from a base or parent class.

   To favor composition over inheritance is a design principle that gives the design higher flexibility. It is more natural to build business-domain classes out of various components than trying to find commonality between them and creating a family tree.

   Initial design is simplified by identifying system object behaviors in separate interfaces instead of creating a hierarchical relationship to distribute behaviors among business-domain classes via inheritance. This approach more easily accommodates future requirements changes that would otherwise require a complete restructuring of business-domain classes in the inheritance model.

G. **ANF Clinical Statements Represent the Minimum Disjoint Set:** Analysis Normal Form (ANF) clinical statements contain the minimum disjoint set of data elements needed to specify a statement using data elements and structures (e.g. topic, result, and circumstance).

H. **Clinical Statement Model Stability:** Stable means that the model can still meet unanticipated requirements without having to change. It is not acceptable to change the model every time a new way to administer a drug or to treat a condition is identified. By representing these types of potentially dynamic concerns in the terminology expressions, as opposed to static fields in a class structure, we do not have to change the model every time something new is discovered. A design imperative is anticipating breakdowns, and providing a space for action when they occur. [11]

   In some regards, in this context “stable” means “not brittle.” A model easily broken by changes that someone could anticipate is one possible definition of brittle. A stable model is critical in the phase of a known changing landscape. We do that by isolating areas of anticipated change into a dynamic data structure. That dynamic data structure may also be immutable in an object that represents a clinical statement.

I. **Reusability:** Architectural patterns should encourage class reusability where possible. Reusability may further refine encapsulation when composition is considered.

J. **No False Dichotomies:** Dichotomies are created when data elements are not mutually-exclusive thus as allowing a certain clinical statement information to be conveyed in two ways. For example if family history were represented in the *topic* field in addition to the *subject of information* field we would characterize it as a dichotomy. False dichotomies lead to arbitrary classification rules and result in ambiguity based on different assumptions about the domain. False dichotomies must be eliminated by ensuring that fields in the model cannot be used interchangeably.

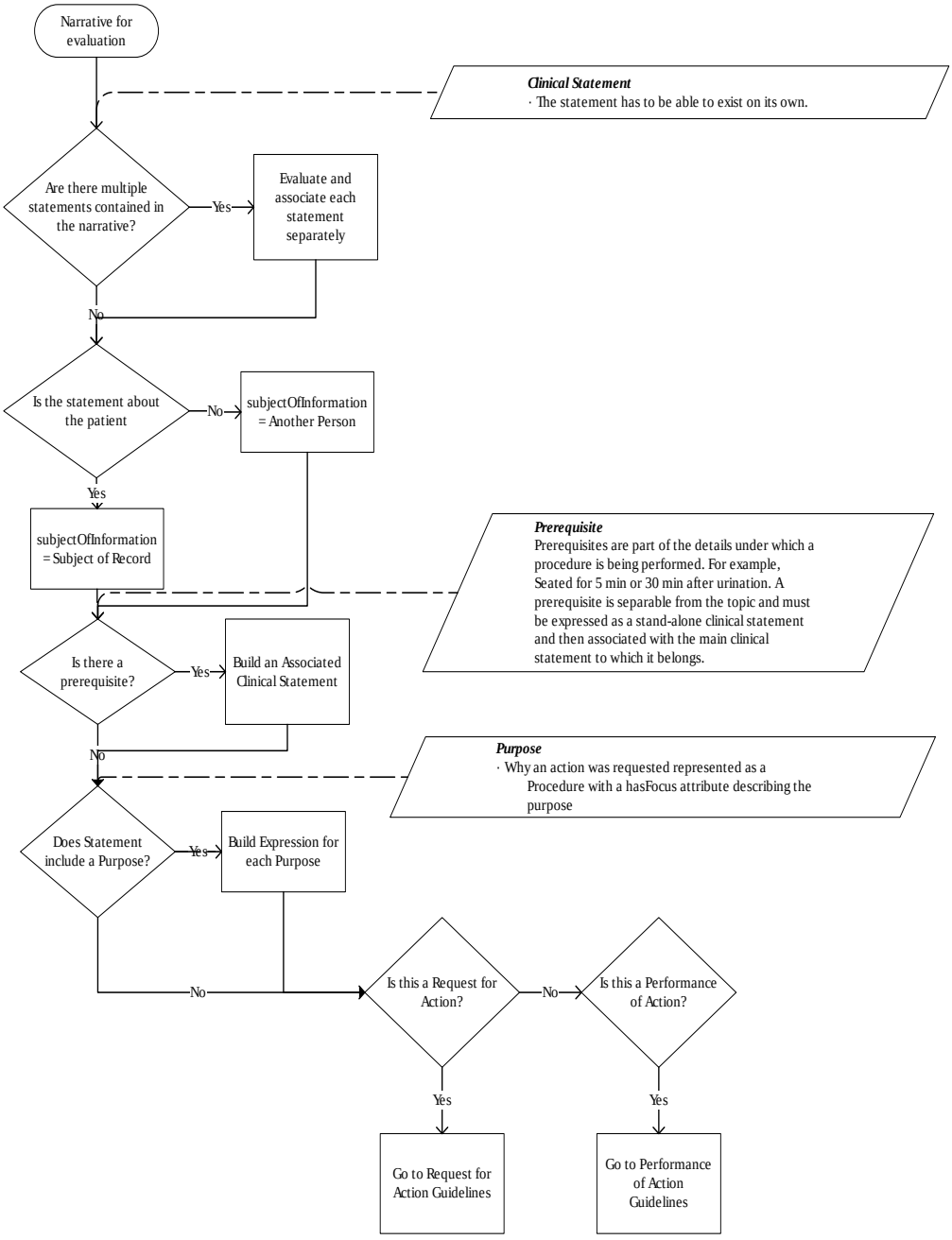
K. **Model Symmetry:** Symmetric models are more consistent, easier to comprehend, and use.

L. **Iterative development and validation of model using use cases:** ANF has been developed using an iterative approach evaluating the model with narrative use cases. Examples of narratives used to evaluate the model can be found in the Appendix.

4.2. Shared Modeling Guidelines

All ANF statements share some common model components. The following modeling guidelines can be used to properly model a narrative into the appropriate components of a single statement or a statement that has multiple associated statements. For the purposes of ANF, a statement is a request for—or performance of—an action that has to be able to exist on its own. Therefore a narrative would be separated into multiple clinical statements if it contains multiple requests or performance of actions that could exist on their own.

Figure 4.1. Shared Modeling Guideline Decision Tree



Editorial Rule: Techniques are inseparable from the topic

- A technique must be true within the duration of the performance.
- A technique is inseparable from the topic and cannot be expressed as a stand-alone clinical statement.

- A technique is a device used, a method applied, or a temporary state in which the patient was actively placed during performance of the action.

Editorial Rule: Prerequisites must be separated from the topic

- A prerequisite must be separable from the topic and should be expressed as a stand-alone clinical statement.

- A prerequisite is a state that must exist before something else can happen or be done. Prerequisites are part of the details under which a procedure is being performed. The state must exist prior to the performance of the action.

Editorial Rule: Subject of information is used to represent family and donor history

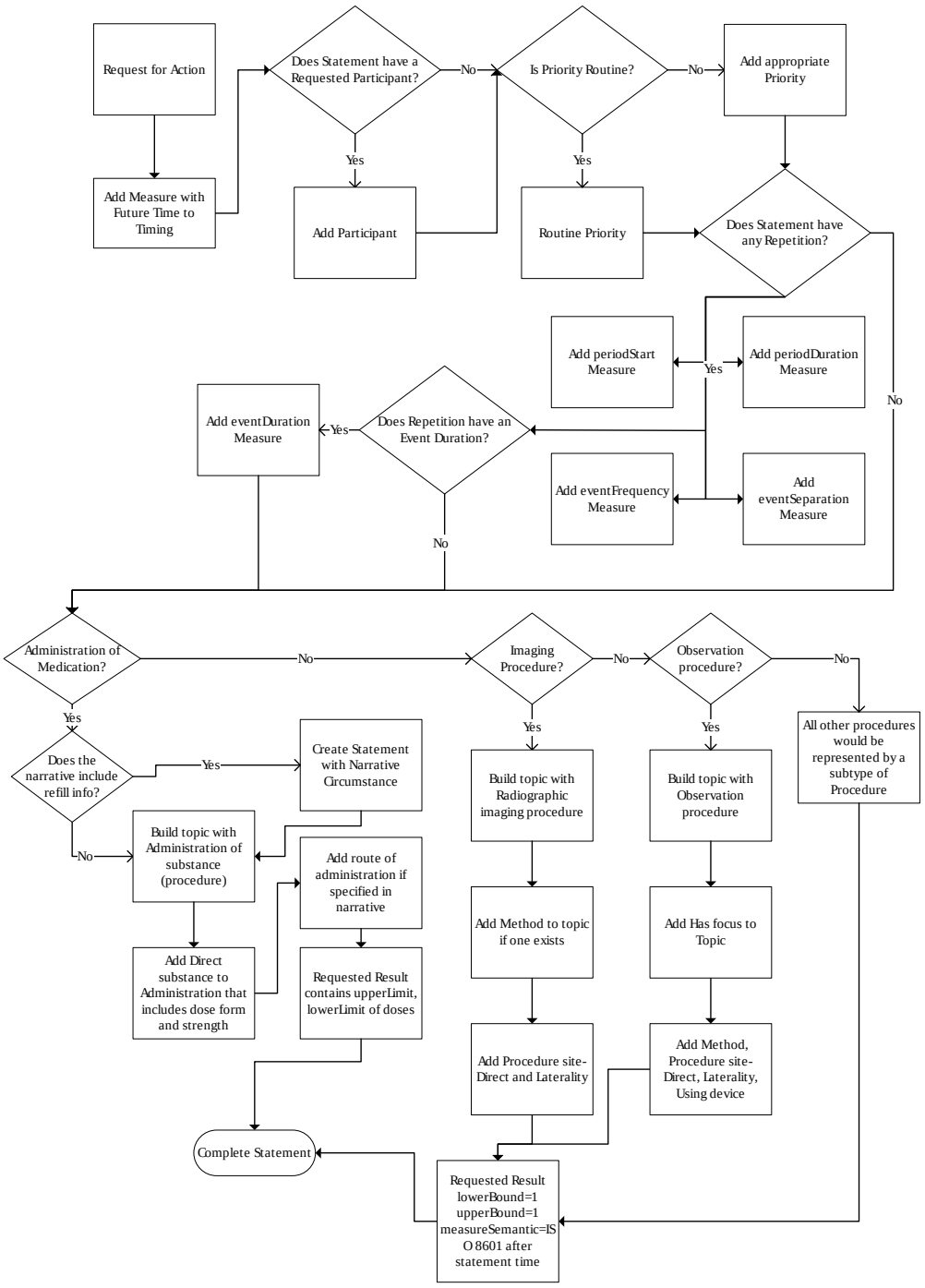
- The subjectOfInformation is used to represent who the statement is about. This is normally the patient (Subject of record) unless explicitly stated otherwise, for example Mother, Sibling, Donor, etc.

Editorial Rule: Purpose indicates the reason for a request or performance

- The purpose is why an action was requested or performed. The purpose of the topic is typically some type of therapeutic intent, diagnostic intent, or both. There can be more than one therapeutic intent and diagnostic intent. While the purpose can also exist as a separate clinical statement, if you specifically want to state that a action was performed for a particular purpose, it must be represented using the purpose.

4.3. Request for Action Guidelines

Figure 4.2. Request for Action Modeling Guideline Decision Tree



Editorial Rule: Timing - past, present, or future

- For a Performance of Action, the Timing can represent a time in the past or a current time. If a history of a performance of action is to be represented in ANF the Timing will be for a time in the past prior to the statement. Otherwise the Timing will be represented with the current time of the statement.

- For a Request of Action, the Timing will always represent a future time.

Editorial Rule: Participants can be specified or requested

- A Performance of action can specify participants using participant in PerformanceCircumstance.

- A Request for action can specify requested participants using requestedParticipant in RequestCircumstance.

Editorial Rule: Priority defaults to routine for a request

- Priority is used to represent the priority for which a request is to be carried out. By default a Request will be considered "routine" unless otherwise specified.

Editorial Rule: Topics are always an action

- The particulars of how topics—and other statement fields—are modeled as a Terminology Layer concern, not a Statement Layer concern. The Statement Layer does require that the Terminology Expression fields in a statement are disjoint: There should be no confusion—or creation of false dichotomies. There should be one, and only one, place to put each type of information in a terminology expression. For example, the Statement Layer defines a particular place to represent the subject of information. Therefore, the Terminology Layer must not allow the subject of information to be redundantly—and possibly contradictory—represented in a topic expression (such as would be the case if "maternal history of diabetes" were an allowed topic expression). The Statement Layer requires that the topic represent an Action as a code or expression according to the rules of the Terminology Layer, and that the rules of the Terminology Layer enforce a disjointness between different types of terminology expressions. Here we present a starting point for what the Terminology Layer editorial rules may look like, based on current SNOMED CT practice.

- SNOMED CT can accommodate this requirement for simple observations by using Observation procedure to represent the topic (or other types of procedures when appropriate, such as the administration of a medication). In SNOMED CT examples, the Observation procedure specifies a Has focus attribute linking it to the Clinical Finding or Disorder that it is being observed. The observation procedure can also be further refined by adding attributes in the terminology model, including Method, Procedure site - Direct, (if appropriate) Laterality, and Using device.

- Medication administrations will use an Administration of substance concept to represent the topic. All Administration of substance concepts will be refined with the substance, dose form and strength being requested. If Route of administration exists, then it will also be added.

- Laboratory tests will use a Laboratory Procedure concept to represent the topic. These concepts can be further refined.

- Imaging Procedures will use an Imaging Procedure concept to represent the topic. These concepts will be further refined with a Method, Procedure site and (if appropriate) a laterality for those sites that are lateralizable.

Editorial Rule: Repetition is used to request multiple occurences of a topic

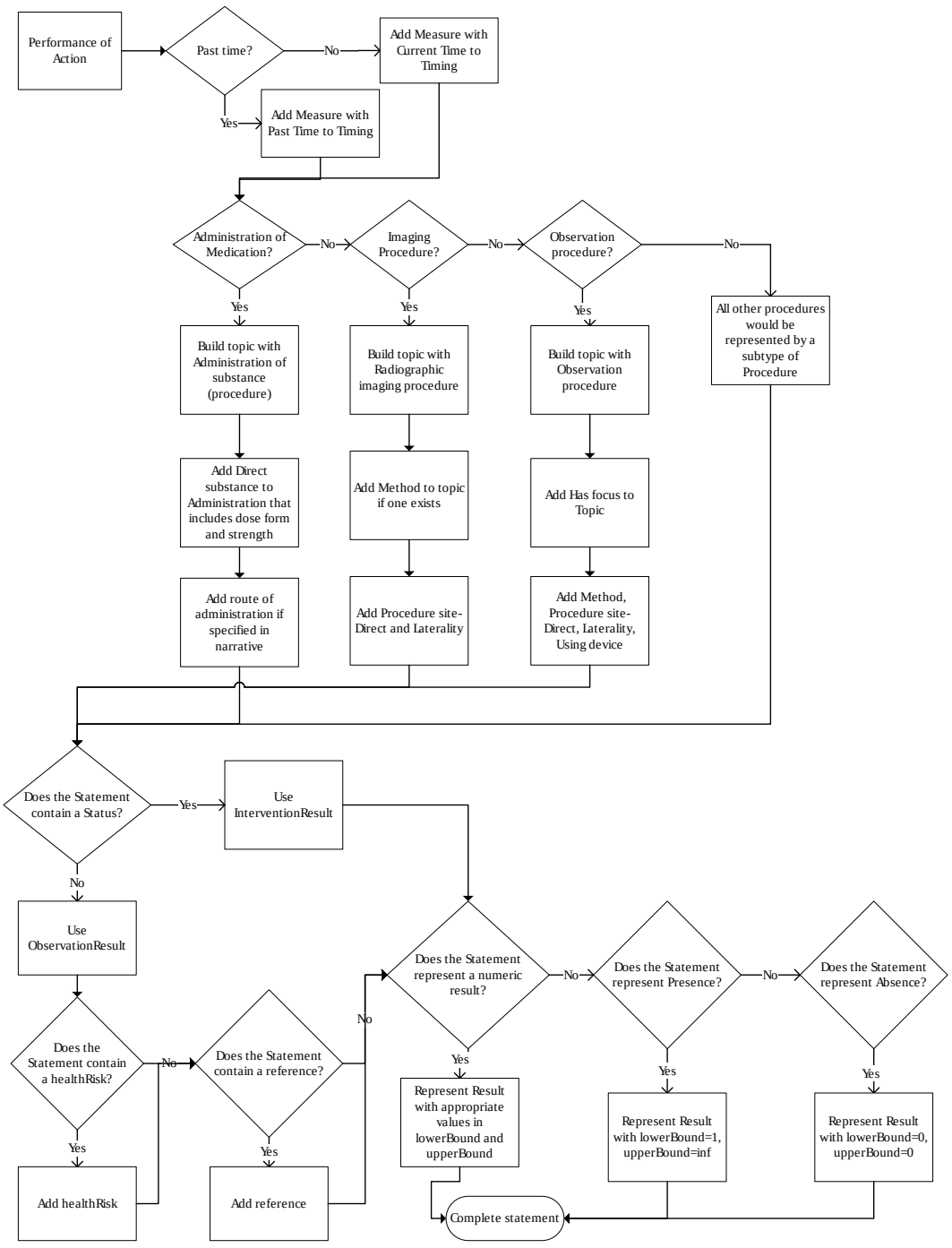
- Repetition is used to represent when an action is requested for more than a single occurrence.

Editorial Rule: A desired result can be specified in a request

- A desired result can be specified as a Measure using requestedResult in RequestCircumstance.
- If a requestedResult is specified, the appropriate upperBound and lowerBound is specified with the correct result semantic.
- If a requestedResult is unspecifiedt, the value is set to [0, inf) with a result semantic of "Countable quantity".
-

4.4. Performance of Action Guidelines

Figure 4.3. Performance of Action Modeling Guideline Decision Tree



Editorial Rule: Timing - past, present, or future

- For a Performance of Action, the Timing can represent a time in the past or a current time. If a history of a performance of action is to be represented in ANF the Timing will be for a time in the past prior to the statement. Otherwise the Timing will be represented with the current time of the statement.

- For a Request of Action, the Timing will always represent a future time.

Editorial Rule: Topics are always an action

Editorial Rule: Status indicates the state of a result

- The status of a Performance of action can be specified with concepts such as "on hold", "completed", "rejected", etc.

Editorial Rule: HealthRisk indicates the clinical risk of the result

- In PerformanceCircumstance, healthRisk is used to flag a result with coded values such as 'low', 'normal', high', and 'critical'.

Editorial Rule: Reference Range can be specified for a result

- In Performance Circumstance "referenceRange" is the interval of values that are normal for the observation/finding described by the "topic" for this "subject". It refers to "normal" for the patient/subject under specific conditions.

Editorial Rule: Results are always a ranged quantity

- Results are always a Measure, which is a ranged quantity. Measure includes both a numeric interval along with a Measure Semantic specified as a Logical Expression.

- If a Result is intended to represent a numeric result then the upperBound and lowerBound would be populated with the appropriate numeric values and the Measure Semantic would indicate the unit of measure.

5. Putting it Together: Normalization and Transformation

Normalization of *clinical statements* is defined as "the ability to identify every representational format that confers the same meaning as being equivalent (i.e., unambiguous representation)." [12]

5.1. Data Structures

Currently, the standard is to define detailed clinical models using different data structures for different domains of clinical statements. For example, *FHIR* independently defines the resources for Conditions, Observations, Diagnosis, Procedure, Goal, Medication Administration, Medication Request, etc. Some implementations, such as FHIR, explicitly define the property names for the parts of each data structure tree and other formalisms such as Basic Meta Model (BMM), Archetype Definition Language (ADL), and Clinical Element Modeling Language (CEML) use a form of key-value pairing to genericise the property naming of the data structure tree. But in all these cases, the fact remains that the resulting structure of the tree still remains different for different domains of clinical statements. Thus, computation and analysis of data instances, that conform to these models, requires a prior understanding of the tree structure for each domain.

ANF seeks to simplify the complexity that currently exists in detailed clinical models. As its name suggests, Analysis Normal Form provides one normalized data structure to describe clinical statements from all domains. ANF accomplishes this simplification by moving the complexity from static statement data structures to dynamic *pre-coordinated*, or *post-coordinated* terminology expressions, as defined by the Terminology Knowledge layer of the architecture.

5.2. Modeling Style

Another variation that currently exists is the allowed design choices which can be made by model authors. For example, a modeler may choose to model breath sounds as 'breath sounds' with a coded result of 'rales', or as 'rales' with a result of 'present'. Currently, organizations try to minimize this type of variation by documenting design choice rules in modeling "style guides". For instance, a common style guide choice in the *CIMI Clinical Statement* model is to either use the *Assertion* style or the *Evaluation Result* style, and *CIMI* documents which types of clinical statement are best suited for each. Assignment of clinical statement types into these categories creates false dichotomies, since there are a myriad of examples where clinical statement types can readily fit in both categories.

ANF's approach is to solve the problem by eliminating the need to make choices between overlapping statement types. ANF seeks to minimize this variation by only allowing quantitative results. This eliminates the choice between Evaluation style versus Assertion style clinical statements as coded results are not possible.

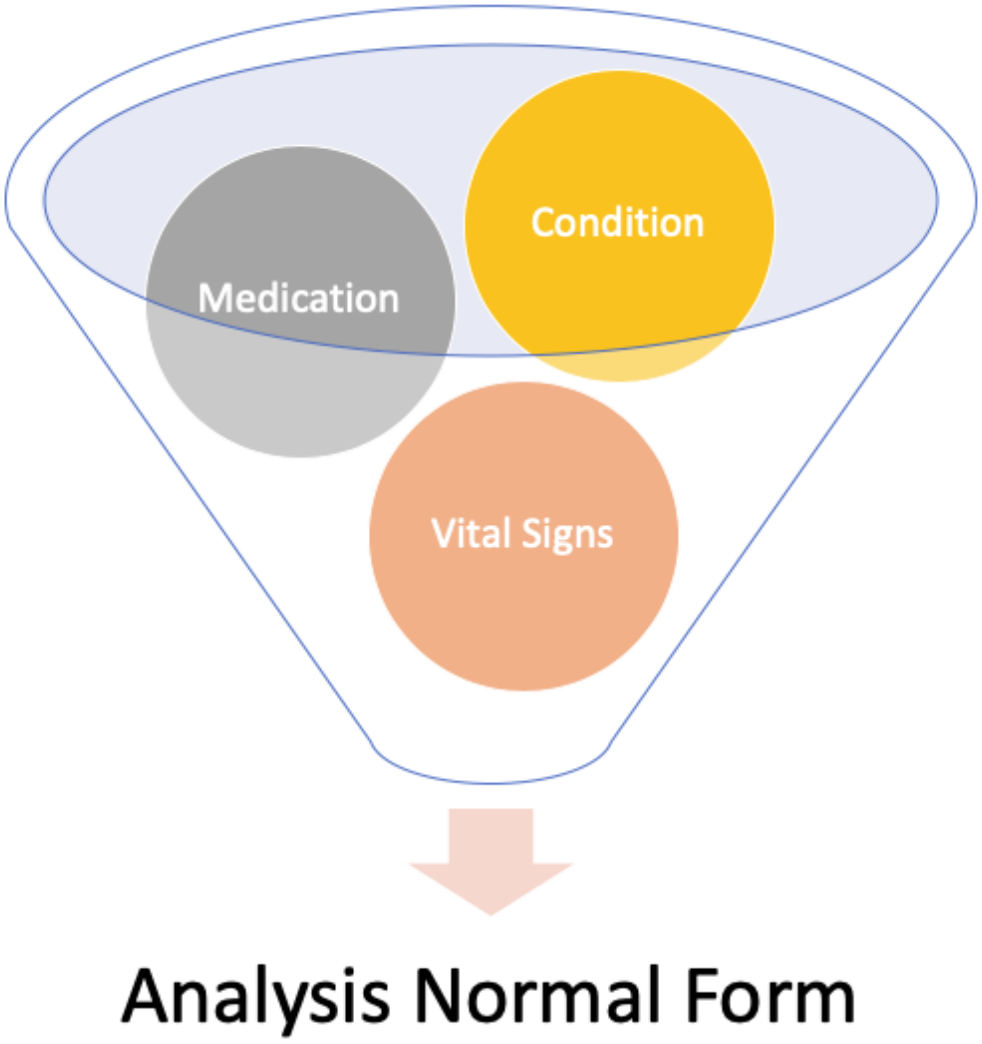
5.3. Transformation to ANF

The previous sections have described the variation that can exist in the data structure and modeling style of a single standard. Moreover, this variation is significantly compounded when simultaneously using data from multiple standards, such as when receiving data from multiple institutions.

Analysis Normal Form can act as a transformation target to normalize these disparate representations of clinical statements, shown in Figure 5.1. Normalization implies the ability to recognize all representations that express the same meaning as being algorithmically equivalent.

To be clear, the transformation discussed is a data instance to data instance transformation. An example could be John Doe's Systolic Blood Pressure measurement taken on June 4, 2019 represented as a FHIR Observation instance, which is then transformed to an ANF instance representing this same data. This is not to be confused with a detailed clinical model transformation between two formalisms, such as an *ISO DCM* for Systolic Blood Pressure transformed to a FHIR profile for Systolic Blood Pressure.

Figure 5.1. Transformation to ANF



Various isosemantic representations of statement models can be brought together into Analysis Normal Form

Transformation, in this case, is not a simple endeavor that one can hope to automate across domains of clinical statements or even within a single domain of clinical statements. As presented, it will involve navigating disparate data structure trees and include variable representations to then generate a well-formed terminology expression. It is most likely possible to target sub-domains for consistent transformation, such as all quantitative laboratory results, but in some cases, it may be that each detailed clinical model needs its own unique transformation.

Potential areas of difficulty during transformation:

- One complex statement/assessment may be transformed frequently be transformed to many ANF instances (see ???)
- Implied clinical meaning or mapping associated with data structures and bound terminology must be transformed into post-coordinated SNOMED CT expression for inclusion as *ANF Topic* (see ??? as examples using explicit and implied semantics to normalize clinical statements.)

Currently, there are three basic categories of errors that might be associated with attempts at normalizing clinical statement representation:

- errors associated with normalization of content of the terminology;
- errors associated with normalization of the semantics of the terminology;
- errors that result from ambiguous or misleading interaction between the structured clinical input and presentation of compound terminology to clinician end-users.

5.4. Transformation Languages and Architecture

A number of options exist for expressing transformation logic and for executing the transformation on specific instances of clinical data for normalization into ANF. These range from transformation languages to expensive middleware options commonly used in healthcare interfaces. The suitability of the chosen language highly depends on the format of the source data, and the quality and accuracy of the transformation is left to the transformation author. One option described here is Model Driven Message Interoperability (MDMI), which is an architecture for transformation that assists in producing semantically accurate transformations.

5.4.1. XSLT

XSLT (eXtensible Stylesheet Language Transformations) is a World Wide Web Consortium (W3C) standard language for the transformation of structured data. [13] XSLT transformation scripts take as input any valid XML document and produce as output an ASCII-formatted document (including XML, HTML, other formatting languages, free text, etc.). The XSLT language specifies transformations through declarative, rule-based commands (see below).

XSLT is widely used in modern information processing, including in health care applications. Numerous XSLT transformation engines exist, including commercial and open-source versions. These implementations are mature, stable, and high-performance, and are available as runtime libraries or embedded in XLST authoring/editing applications. Excellent documentation and training are available for XSLT.

XSLT scripts operate over source “trees” containing the structured contents of parsed XML documents. These trees contain as their nodes the various constructs of specific XML documents, i.e., the named elements, attributes, and text values that appear in the documents, and upon parsing, becomes a source tree for XSLT transformations.

XSLT uses the sub-language “XPath” to reference portions of the XML source tree for purposes of navigating the tree and selecting specific parts of it to translate. [14] XPath is essentially a query language for identifying and retrieving XML sub-trees that match specified criteria.

The actual transformation logic in XLST scripts is specified as a series of “templates”. Each template matches to a specified sub-part of the source tree and specifies what output will be generated for that sub-part. Templates are generally called from within other templates via a declarative template-matching

process, and a recursive traversal and transformation of the input tree occurs through this template-invocation model. The transformation logic within templates may include various conditional, branching, and formatting constructs, as well as calls to external functions written in various programming languages (such as Java).

XSLT is effective in representing and executing the transformation logic needed for clinical translations. In general, XSLT provides various advantages, as well as limitations, for this task.

Advantages

- A powerful language
- Declarative – automated matching of templates to data
- Extensible via extension functions and external function calls
- Many mature implementations
- Good tooling (e.g., Eclipse plugin, XMLSpy)
- Good documentation

Limitations

- Transformation specifications are verbose and hard to read/understand/debug/maintain
- Transformations are entirely syntactic
- Limited to XML input – instances rendered in other formats cannot be translated

5.4.2. FHIR Mapping Language

The FHIR mapping language (FML) [15] is a relatively new, bespoke transformation language specifically designed to transform HL7 FHIR resources to alternative representations, including different FHIR resources, C/CDE documents, etc. The mapping language was created by the FHIR Management Group as a specification of the QVT framework for model-transformation languages (see Section 5.4.3, “QVT”).

Conceptually, FML is similar to XSLT in that it (a) consists of declarative rules that are automatically matched to input data, (b) includes a sub-language (“FHIRPath”) to reference parts of source parse trees, and (c) has the ability to reference external functions written in different languages. There are also notable differences between FML and XSLT. The source input of FML is not limited to XML documents, but may include any object models and rendering syntaxes conformant with *OMG’s Meta Object Facility (MOF)* language. [16] MOF is a general formalism for representing object models as directed acyclic graphs (DAGs), and MOF-compliant models can use various syntactic constructs to represent the classes, attributes, and attribute values of such graphs.

Hence, in FML, there is no built-in notion of source trees containing XML “elements”, “attributes”, “comments”, “namespaces”, etc. In fact, FML transformation rules do not specify any target syntax for inputs or outputs, just the general concepts of named classes, class members, and member values. This flexibility would allow transformation source inputs used in the normalization to ANF to be represented in different formats than XML, were that to be deemed preferable. For example, instances rendered using JSON, ODIN, or ASN1 syntax could be the inputs of FML transformations.

The output of an FML transformation is not a text-rendered document (unlike XSLT), but an internally stored DAG consistent with the specified output model. Subsequently, the DAG may be rendered in any number of syntaxes, including XML, JSON, or the tables and fields of a relational database.

The FHIR Mapping Language may also be effective in representing and executing the transformation logic needed for normalization to ANF. As with XSLT, however, there exist certain trade-offs in its use.

Advantages

- Support for input formats other than XML
- Transformation logic produces semantic DAGs, which can be subsequently rendered in a variety of syntaxes.
- The mapping specifications are more concise and easier to read/understand than XLST

Limitations

- Inputs/outputs other than FHIR *logical models* currently require additional custom programming
- Only XML and JSON are currently supported as output syntaxes without custom programming
- Only one implementation to date (as a library)
- Limited tools for authoring/editing transformation scripts
- Limited sources of documentation
- Few knowledgeable programmers

5.4.3. QVT

A third alternative is to develop a new transformation language customized to support the requirements of a normalization to ANF, based on the QVT language used to develop the FHIR Mapping Language.

QVT [17] is a general model-transformation framework and language developed by the Object Management Group. It includes both an imperative ("QVT-O") and a declarative ("QVT-R") version, and offers considerable flexibility in defining the constructs of purpose-specific transformation languages. Although QVT is intended for the transformation of data *models* rather than data instances, the FHIR Mapping Language shows that it can be applied to the latter task as well.

A number of implementations of QVT exist as open-source and commercial software offerings. These include:

- ATL (open source). Probably the most widely used and maintained of the available implementations. Includes a library of existing QVT transformations, to serve as examples and templates.
- Eclipse M2M Project (open source). An Eclipse project that includes authoring tools for QVT transformations, as well as various transformation engines (including the one from ATL).
- ModelMorf (proprietary)
- Others (see [17])

Advantages

- QVT is very abstract, which confers great flexibility and configurability to create custom transformation languages.

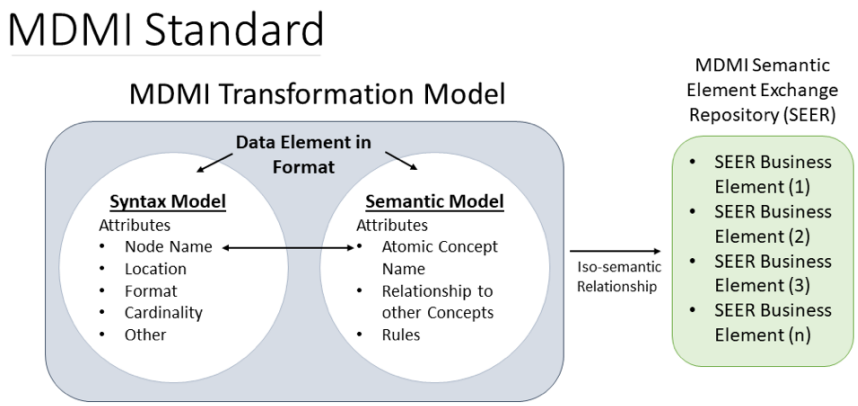
Limitations

- The abstractness also makes QVT quite difficult to understand and learn, and there are limited resources to assist in the learning process.

5.4.4. Model Driven Message Interoperability (MDMI)

MDMI is an Object Management Group Standard for the transformation of data in one format to data in another format. MDMI Standard is not a language. The MDMI Standard is a specification for addressing this problem and was developed by multiple domain experts. The specification contains two major sections: the MDMI Transformation Metamodel and the MDMI Semantic Element Exchange Repository (SEER).

Figure 5.2. MDMI Standard



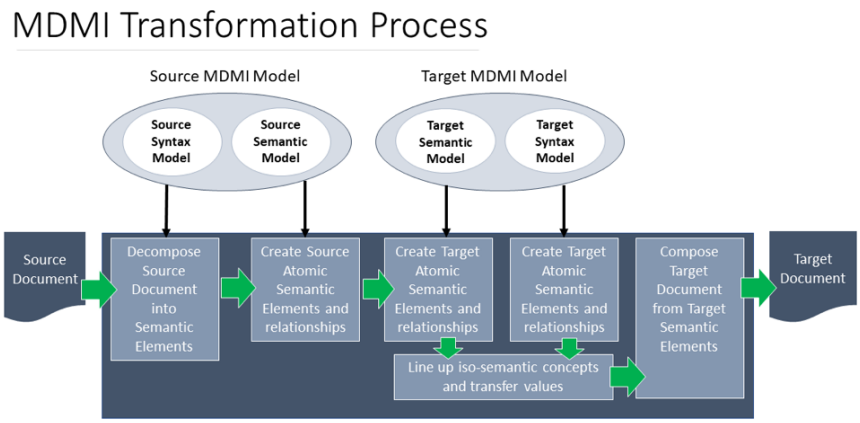
The MDMI Transformation Metamodel.

The MDMI Transformation Metamodel is composed of a syntax model and a semantic model. The syntax model contains the syntactical representation of each data element in a format and the semantic model contains the semantic concept represented by the data element. The syntax model is used to compose a collection of semantic representations into a target file format or to decompose a source file into its semantic representations. The syntax model can be used for any format. XML, JSON, HL7 V.2, CVS, various EDI payment, and proprietary formats have been used.

The semantic model captures the semantic concepts in the format and the relationships between the semantic concepts in a format. Probably the most important relationship is the containership relationships. The semantic model of the MDMI is also used to capture other relationships and rules required to create unambiguous semantic representations. An example of this is a data element that can have multiple semantics concepts that must be disambiguated based on other values contained in format.

The MDMI SEER is a repository for the semantically unique concepts, called Business Elements, that are exchanged in healthcare transformations. One can view the MDMI SEER as a bag of unique, atomic semantic concepts exchanged, primarily driven by the *HL7 standards of v.2*, *CDA*, and FHIR that are used to exchange information. If there is a new semantic concept that does not exist in the SEER, then a new Business Element is simply added. Each MDMI Transformation Model uses the MDMI SEER to create an iso-semantic relationship with its own semantic concepts and a Business Element.

Figure 5.3. MDMI Transformation Process



The MDMI Transformation Process

There is a project underway in the OMG to extend the MDMI SEER. The Business Elements in the MDMI SEER are *pre-coordinated semantic concepts* represented in industry standard healthcare ontologies and terminologies. The project is using the ANF Statement Model as a Reference Model to develop a semantic model that can precisely define the meaning of the Business Element in a detailed, structured, unambiguous, computable formalism.

An open source implementation of MDMI started in the Open Healthcare Tools organization which built an MDMI compliant tooling for healthcare. The MDMI Open Source Project continues today in GitHub and has been and is being used in HL7 projects as well as in commercial implementations.

MDMI is a model driven approach. Having a formal model, the open source project has been able to develop tooling based on the MDMI model as well as leverage other modeling efforts. Examples are Information Models such as FHIR and the FHIM using the model driven MDHT tooling and Ontological Models such as ANF / *Solor*.

Advantages

- Any-to-Any transformations versus point-to-point language mappings allow reuse of transformation models for different use cases.
- It minimizes change. If one MDMI Model changes (e.g. FHIR 4 to FHIR 5), this does not require changes to other existing MDMI Models such as CCDA 2.1, HL7 V2.8, or a proprietary model.
- It simplifies development. Tooling exists to develop and maintain individual MDMI Models by SMEs who do not need to be developers. The scope of expertise is further reduced because the knowledge one needs to create a MDMI Model is primary to know what the data in their format means.
- It enables automation tooling for creating MDMI models, for creating computable artifacts, and generating reports.
- There are Open Source Models for HL7 formats as well as the MDMI tooling.

Limitations

- MDMI has limited experience with transformations of detailed clinical models.
- User Documentation of MDMI is lacking.
- The MDMI runtime tool is complex.

6. Pragmatic Usage and Next Steps

Like other *CIMI isosemantic models*, *ANF* is a *logical model* and therefore it may be implemented using relevant implementable models and technology (see SAIF-CD). Thus, this project will expand on the use of ANF alongside preexisting information exchange HL7 standards (i.e. *HL7 V2 messaging*, *CDA documents*) and HL7 standards-based APIs (i.e., FHIR resources). In practice, ANF is applicable to systems normalizing or creating normalized data to support *Assertional* and *Procedural* knowledge (e.g. clinical alerts, workflow, data analysis, decision support). These need to aggregate data from many isosemantic source models into a single analysis format.

Implementers may use the logical model and methodology in this document to design software components, databases, or APIs that support reuse and analysis of treatment information captured using best practices (e.g. *CIMI models*) and exchanged using interoperability specifications required across the US (e.g. FHIR US Core, Consolidated CDA). Since information sharing already relies on a variety of clinical statement approaches and syntax representations, it will become necessary to create normalized instances of those clinical statements intended for reuse. Not all the data produced by a system is necessary for analysis; and, the ANF model—like other CIMI models—is focused on clinical information. ANF does not require a specific input form syntax; its focus is on implementations where data quality and semantics and on implementations that require information must be aggregated from many and diverse sources.

The ANF logical model can be used to create practical implementation guidance (i.e. implementation guide, profiles, value sets based on standard terminology) and can be applied to design data analysis solutions. Implementation specifications include vocabulary bindings based on standard terminologies (e.g. SNOMED CT, LOINC, RxNorm) to support the Terminology layer of the *Knowledge Architecture*. For simplicity, SNOMED CT is used for all logical expressions and examples in this specification but ANF implementation may require LOINC, RxNorm, or other standard terminology.

6.1. ANF FHIR implementation

ANF is a logical model intended to represent any clinical data using a complete yet simple normal form. It allows other software modules to reuse the information and derive new knowledge from it. Examples of ANF's benefits include improved ability to (1) analyze the care that was delivered, (2) find out what type care leads to the best patient outcomes, and (3) use rules and business triggers to automate clinical decision and workflow steps. ANF could be used to design standards-based Application Programming Interfaces (APIs) optimized for a specific analysis purpose. ANF APIs may be implemented using FHIR resources, profiles, and extensions to access clinical decision support, clinical quality measures, and to support workflow automation by triggering reminders and clinical notifications.

ANF statements may be created from existing clinical statements and patient-entered data to support APIs intended for analysis or to automate information derived from device measurements, clinician inputs, and patient-generated data.

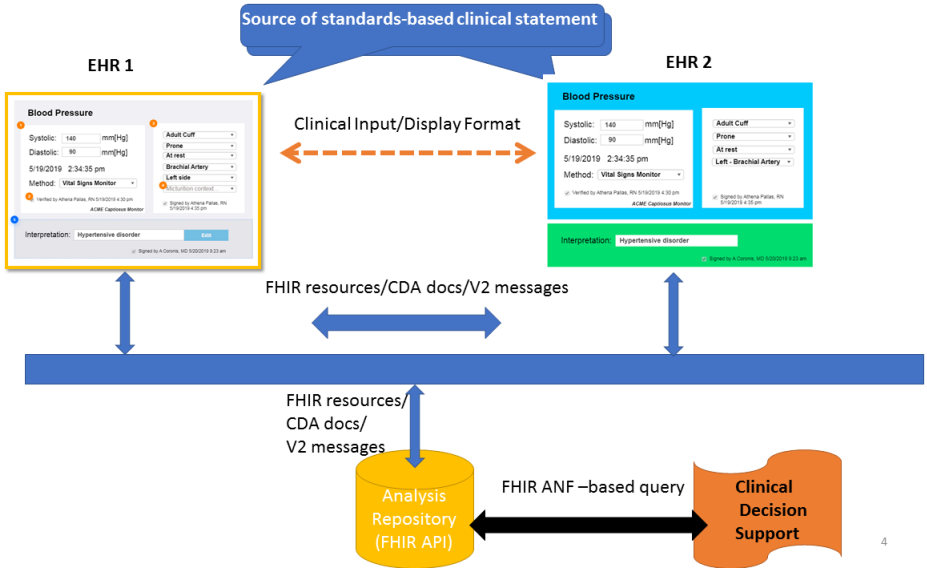
6.1.1. Analysis API

The typical implementation of ANF will be a system that normalizes clinical information (e.g. FHIR, CDA documents) to be used by business and decision support rules. Healthcare enterprises may use middleware, standards-based transformation and terminology servers to normalize a variety of observations, orders, diagnoses, medications, procedure notes, and other interventions to a set of *Performance* or *Request* statements. Narrative clinical statements may not be immediately reducible to ANF and it may require natural language processing and other methods of augmentation and enhancement.

ANF-specific resources and implementation guidance can be tested during FHIR Connectathons to validate that the logical model outlined in this specification is suitable to data aggregation and supports the analysis

objectives of researchers, before proposing them as new resources for future versions of FHIR and as extensions and profiles for current versions. Both approaches may be desirable.

Figure 6.1. ANF-based FHIR API

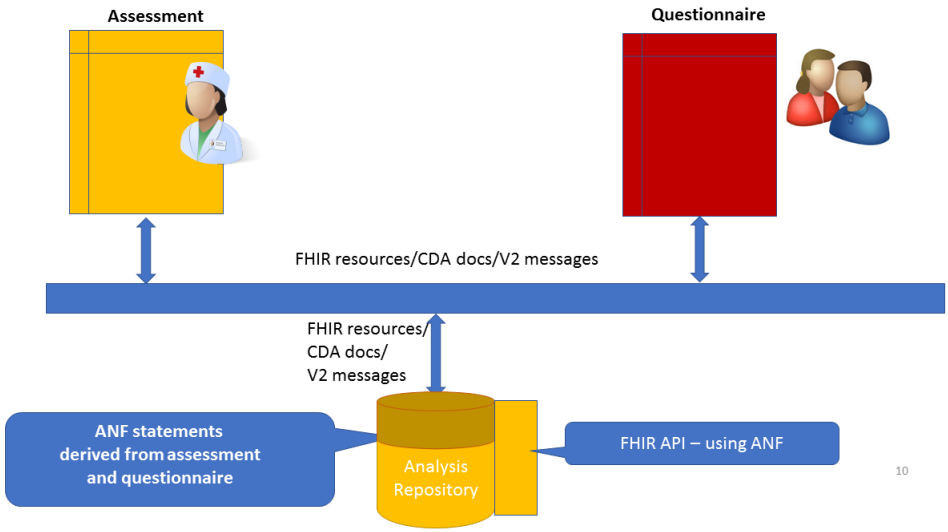


ANF-based information can be used to create data warehouses and support data mining.

6.1.2. Automated Data Analysis

ANF statements may be created as an outcome of evaluating device, clinician, and patient-entered data (e.g. questionnaires) automatically and in near-real-time. For example, specific answers to a PHQ-9 screening tool along with previous assessments could trigger a specific type of follow-up screening regarding substance use treatment or further evaluation, consideration of *Social Determinants of Health* (SDOH), or alert to a provider. While ANF statements are not intended as an input form, such statements could be automatically generated by Learning Health Systems [18] using a combination of pre-existing clinical data, clinical guidelines/rules, medical device observation and patient-generated data. The promise of the *Learning Health System* [18] is the ability to learn new knowledge from previous clinical statements and latest scientific developments. This approach is also conducive to tailoring treatment consistent with *precision medicine* [19] and reducing provider burden through automation.

Figure 6.2. Deriving ANF Statements

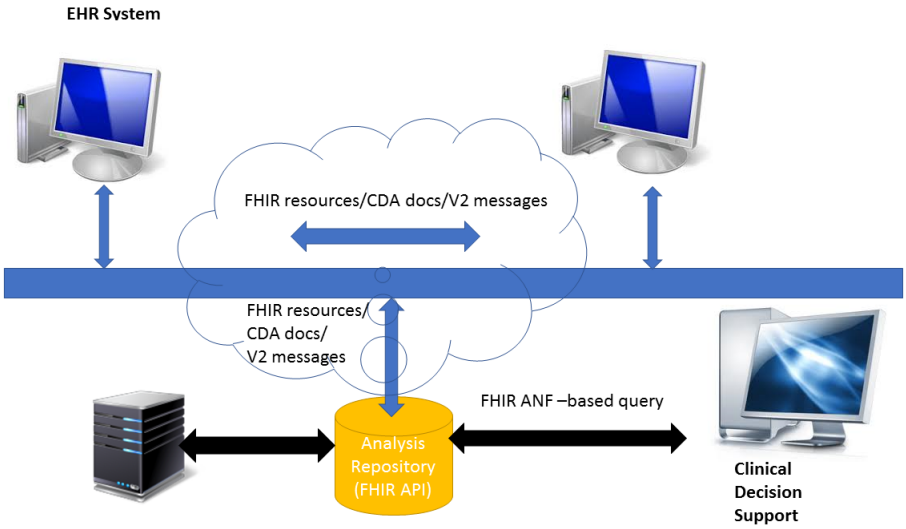


ANF statements may be derived from other data inputs and combined to support near-real-time analysis.

6.2. Other platforms

Big data analytics, data mining, business intelligence, healthcare quality programs, registries etc. all require large data sets of consistent structure and semantics that can be analyzed and aggregated for the benefit of individual patients, to evaluate an organization, or to establish new facts.

Figure 6.3. Data Mining using ANF statements



Standards-based information may be normalized to ANF to be used for clinical decision support.

Interoperability standards sometimes pose challenges due to the use of synthetics/abstract data types that attempt to capture the complexity of healthcare data. ANF simplifies the statement structure by using a

small set of primitive types (e.g. float, varchar, boolean) and a sophisticated terminology. Data warehousing and mining solutions rely on a consistent simple representation of data organized along facts and axes. ANF borrows from database normalization the idea that "*normalization*" reduces data redundancy and improves data integrity. The ANF logical model can be used to design "fact-based" dimensional schemas for databases which enable analysis of a specific set of facts and dimensions, such as evaluation of outcomes associated with the use of a specific therapy, device, or medication.

7. Implications—Improving Patient Safety and Outcomes

ANF has implications on clinical data quality, *clinical decision support*, patient safety and population health because it promotes the reuse of information aggregated to derive new information about treatment quality, patient safety, and outcomes.

7.1. Improved Data Quality

Information systems record and manage clinical statements using a variety of standard or ad-hoc models and formats. However, analysis of clinical statements requires consistency, not only at the format level (e.g. *CDA*, FHIR, *V2*), but also at the content and semantic levels (i.e. ANF, *CIMI model*, etc.). In most cases, the data quality is the greatest obstacle to analysis. Analysis Normal Form aims to minimize data quality challenges and provide a common format with semantic clarity to allow for meaningful secondary uses of clinical data.

The design of ANF is based on research into data quality frameworks [20] which identified that information conformance, completeness, and plausibility are all necessary to analysis.

- **Conformance**: Conformance describes how well a system or implementation meets a specification. ANF provides a logical structure and constraints of clinical data for value conformance, relational conformance, and computational conformance irrespective of data representation (e.g. CDA, FHIR).

    - Value Conformance: Value conformance seeks to determine if recorded data elements are in agreement with a predefined, constraint-driven data architecture. Internal data constraints are typically imposed by the *ANF Reference Model*.

    - Relational Conformance: Relational conformance seeks to determine if the recorded data elements are in agreement with additional information referenced by a *clinical statement*. An *ANF Statement* may reference other information about patients, practitioners, encounters, etc. to provide context to the topic and result recorded.

    - Computational Conformance: Computational conformance seeks to determine if computations used to create derived values from existing variables yield the intended results either within a data set (Verification) or between data sets (Validation), when programs are based on identical specifications. Computational conformance focuses on the correctness of the output value of calculations against technical functional specifications. ANF highlights the measure in which an action, finding, or observation was either requested or performed to a common "measure" thus supporting the development of computational, assertional, and procedural predicates.

- **Completeness**: Completeness focuses on features that describe the frequencies of data attributes present in a data set without reference to data values. Completeness measures assess the absence of data at a single moment over time or when measured at multiple moments over time. ANF disambiguates the date a statement was made/asserted from the timing of the circumstances in which the underlying action, observation, or finding occurred.

- **Plausibility:** Plausibility focuses on features that describe the believability or truthfulness of data values. For this category, plausibility is determined by a variable’s value, when a value is placed within the context of another variable (i.e., two independent variables assessing the same construct), or a temporal sequence or state transition (e.g., patient follow-up treatment for a disease must be preceded by a corresponding diagnosis).

- Uniqueness Plausibility**:** The Uniqueness subcategory seeks to determine if objects (entities, observations, facts) appear multiple times in settings where they should not be duplicated or cannot be distinguished within a database (Verification) or when compared with an external reference (Validation). Duplication frequently occurs when disparate data streams that contain overlapping objects are combined. ANF provides the contextual data needed to de-duplicate clinical statement prior to analysis.

- Atemporal Plausibility: Atemporal Plausibility seeks to determine if observed data values, distributions, or densities agree with local or “common” knowledge (Verification) or from comparisons with external sources that are deemed to be trusted or relative gold standards (Validation). For example, in the case of systolic blood pressure, an independent verification of the value measured by a device is provided by the practitioner who conducts performance. ANF clinical statements support results that are evaluated based on a "reference range" of plausible values based on patient status, device-supported ranges, or human physiology.

- Temporal Plausibility: Temporal plausibility seeks to determine if time-varying variables change values as expected based on known temporal properties or across one or more external comparators or gold standards. Temporal properties that establish expectations in this subcategory include temporal stability (do values vary over time as expected), temporal continuity (do values persist over time as expected), state transitions (do sequences of events occur as expected), and temporal dependencies between time-varying variables.

7.2. Enhanced Clinical Decision Support

A 2012 Literature Review commissioned by the Agency for Healthcare Research and Quality (AHRQ) found evidence showing that CDS had a positive impact on process measures and increasing user knowledge relevant to a medical condition. [21]

Additional studies show that well-executed CDS can [21] :

- reduce adverse drug-drug interaction events and medication errors;
- decrease unnecessary laboratory testing;
- reduce cardiovascular risk in patients with type 2 diabetes;
- improve practitioner performance;
- increase cardiovascular disease risk assessment in routine primary care practice;
- improve public health outcomes associated with outbreaks of food-borne illness;
- and, produce cost savings associated with hospital-based pharmacy interventions.

Taken together, the available evidence shows that CDS—when implemented in the right context, and when governed with formal management—can reduce errors, improve the quality of care, reduce cost, and ease the cognitive burden on health care providers.[21] As a result, the impetus for achieving standardized, widespread adoption of CDS across health systems is clear.

A report entitled “*Optimizing Strategies for Clinical Decision Support: Summary of a Meeting Series*” [21] was produced out of the collaboration between the Office of the National Coordinator for Health Information Technology (ONC) and the National Academy of Medicine (NAM). The report states that there are at least four important technical challenges to sharing and therefore standardizing implementations of CDS content: [21]

(1) insufficient standardization of patient data representation;

(2) insufficient standardization of CDS knowledge representation;

(3) insufficient standardization of CDS integration mechanisms;

(4) a need to align with broader standardization initiatives.

One of the reasons that CDS interventions are difficult to implement between health care systems is because disparate EHR systems and health care systems utilize different underlying patient data models and clinical statement representation mechanisms. Even distinct instantiations of use of the same EHR systems differ in how they encode patient data and in how they represent clinical statements. The ONC and NAM report states that "[b]ecause CDS relies on inferencing using patient data, this heterogeneity in patient data representation poses an immense obstacle to sharing CDS." [21]

ANF aims to reduce the variability of clinical data managed by clinical information systems and stored in data repositories. The standardization of clinical observations in a manner that supports automated processing requires a formal clinical statement model, such as ANF. The most important requirements of such a statement model are that (1) it can represent any clinician-specified observation accurately and precisely and (2) it can support automated query and retrieval operations correctly and efficiently.

ANF aims to reduce the variability of clinical data within the value sets and clinical decision rules managed by EHR systems and modeled/stored in data repositories. For example, a clinician could document that a patient has “bacterial pneumonia caused by methicillin-resistant Staph. Aureus” by combining the pre-existing concept “bacterial pneumonia” with the pre-existing concept “Methicillin Resistant Staph. Aureus” and specifying that the latter is the “causative agent” of the former. The patient’s medical record would then contain an entry consisting of the following expression:

Bacterial Pneumonia (ConceptID = 53084003) : Causative Agent (ConceptID=246075003) = Methicillin Resistant Staph. Aureus (ConceptID=115329001)

If specified correctly, post-coordinated expressions also support subsumption testing. Hence, the patient whose record contains the expression above would also be identified by the query “find all patients with a diagnosis of any infectious disease (Infectious Disease : ConceptID = 40733004) in their record.”

7.3. Increasing Population Health

Electronic clinical quality measures (eCQMs) and CDS alerts are triggered by clinical data that is represented in data repositories by clinical statements represented by detailed clinical models with data elements encoded by standards-based clinical terminologies. Because these measures and alerts intend to promote evidence-based clinical processes, variations in clinical data caused by having inaccurate, incomplete, or antiquated implementations of underlying logical models may impact the ability of clinicians to assess care and improve quality. Health information technology-supported quality improvement (QI) initiatives can decrease disparities for some chronic disease management and preventive measures QI. [22] Data-driven QI efforts rely heavily on patient-level data generated by eCQM reports or CDS alerts, which are dependent upon standards-based encoded clinical data. If clinicians rely on inaccurate implementations of eCQMs and CDS, then they may have lists/alerts with patients intended to be excluded from a measure/alert, and may therefore, target inappropriate patients for therapies, such as recommending aspirin use for someone at high-risk for a fatal bleeding event. Similarly, life-saving treatment may be denied or delayed.

Implementation research shows that variations in implementations of eCQM specifications for cardiovascular event prevention could result in potential lives saved or harms avoided in quality improvement activities. [23] For aspirin use for secondary prevention of heart attacks, *Number-Needed-to-Treat (NNT)* statistics show that of patients with known cardiovascular risk who took aspirin, 1.3% were helped by preventing a non-fatal heart attack, and 0.25% were harmed by a major bleeding event. An implementation study [23] against clinical data from two primary care clinics shows that 121 (92%) of the patients were inappropriately included in a measure’s denominator. These patients were also taking an anticoagulant medication, so the *Number-Needed-to-Harm (NNH)* statistic for this subset of patients for aspirin usage is likely much higher, and for this study, 1 to 2 people may have been harmed if the inaccurate implementation persisted, as evidence shows that patients with combinations of aspirin, warfarin, and clopidogrel are

associated with up to a three-fold higher risk of bleeding for patients on dual therapy and triple therapy. With another measure for statin therapy, 1 in 21 people have a repeat heart attack, stroke or death avoided, so even 10 missed people have significant risk of events. Similarly, 10% are harmed by muscle damage or pain, or ~1 of the 14 inappropriately included in the study. [23] Even in the small eCQM implementation study [23] with data from two primary care clinics, failure to include or exclude patients could have led to real harm.

With eCQM implementation and QI infrastructure increasing, the problem of having, and using, inaccurate eCQM implementations or CDS implementations could have significant potential negative impact on population health by not avoiding adverse events and harm to patients. ANF reduces these erroneous implementations. Without a precise logical model for clinical data like ANF, comparability of eCQMs for payment programs and utility of CQM data for targeted quality improvement may be limited.

7.4. Summary

In conclusion, Analysis Normal Form (ANF) presents a simple reproducible approach to modeling clinical statements specifically for data analysis. It reduces clinical statements to two types, Performance of an action, finding, or observation and Request for Action, both clinical statement types with topics. ANF is compatible with other work in statement representation models such as the CIMI Clinical Statement approach, with its focus on more traditional complex structured trees, whereas ANF focuses on structuring data in a way that CDS systems can extract data in an unambiguous way. ANF provides a single, normalized, form for clinical statements that may be used to create assertional or procedural knowledge artifacts, such as clinical decision support rules and clinical alerts.

8. Bibliography

1. Ratwani R, Hettinger Z, Rollin F. The Role of Health IT Developers in Improving Patient Safety in High Reliability Organizations [Internet]. Anticipating Unintended Consequences of Health Information Technology and Health Information Exchange. The Office of the National Coordinator for Health IT; 2014. Available from: https://www.healthit.gov/sites/default/files/medstar_hit_safety_1_29_v2.pdf.

2. Chassin MR, Loeb JM. High-reliability health care: getting there from here. Milbank Q. 2013;91(3):459-90.

3. Spackman KA, Reynoso G. Examining SNOMED from the perspective of Formal Ontological Principles: Some Preliminary Analysis and Observations. Proc. KR-MED, 2004: 72-80.

4. Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. Methods of information in medicine. 1998;37(4-5):394-403.

5. Ahn S, Huff SM, Kim Y, Kalra D. Quality metrics for detailed clinical models. International journal of medical informatics. 2013;82(5):408-17.

6. Handler, J. The Importance of Accurate Blood Pressure Measurement. The Permanente Journal (2009) 13:3: 51-54.

7. O'Brien E, R. Asmar, L Beilin, Y Imai, J. Mallion, G. Mancia, T. Mengden, M. Myers, P. Padfield, P. Palatini, G. Parati, T. Pickering, J. Redon, J. Staessen, G. Stergiou, P. Verdecchia. European Society of Hypertension recommendations for conventional, ambulatory, and home blood pressure measurements. Journal of Hypertension (2003) 21: 821-848.

8. Pickering, T.G., J.E. Hall, L.J. Appel, B.E. Falkner, J. Graves, M.N. Hill, D.W. Jones, T. Kurtz, S.G Sheps, E. J. Roccella. Recommendations for Blood Pressure Measurement in Humans and Experimental Animals: Part 1: Blood Pressure Measurement in Humans: A Statement for Professionals From the Subcommittee of Professional and Public Education of the American Heart Association Council on High Blood Pressure Research. Hypertension (2005) 45:142-161.

9. HL7 Service-Aware Interoperability Framework: Canonical Definition Specification, Release 2. Health Level 7 International.Available from: https://www.hl7.org/implement/standards/product_brief.cfm?product_id=3.

10. HL7 Version 3 Standard: Clinical Statement CMETs Release 1. Health Level Seven International. Available from: https://www.hl7.org/implement/standards/product_brief.cfm?product_id=40.

11. Winograd T, Flores F. Understanding computers and cognition: a new foundation for design. Boston: Addison-Wesley; 2008.

12. Elkin P. Terminology And Terminological Systems. Springer London LTD; 2016.

13. XSL Transformations (XSLT) Version 2.0. Worldwide Web Consortium. Available from: https://www.w3.org/TR/xslt20/.

14. XPath Syntax. Worldwide Web Consortium. Available from: https://www.w3schools.com/xml/xpath_syntax.asp.

15. FHIR Mapping Language. HL7 FHIR. Available from: https://www.hl7.org/fhir/mapping-language.html.

16. MetaObject Facility Specification. OMG MetaObject Facility. Object Management Group. Available from: https://www.omg.org/mof/.

17. Query/View/Transformation Specification.Object Management Group. Available from: https://www.omg.org/spec/QVT/About-QVT/.

18. Friedman C, Rubin J, Brown J, Buntin M, Corn M, Etheredge L, et al. Toward a science of learning systems: a research agenda for the high-functioning Learning Health System. Journal of the American Medical Informatics Association : JAMIA. 2015;22(1):43-50.

19. Precision Medicine in Cancer Treatment [Internet]. National Cancer Institute. [cited 2019Aug1]. Available from: https://www.cancer.gov/about-cancer/treatment/types/precision-medicine.

20. Kahn MG, Callahan TJ, Barnard J, Bauck AE, Brown J, Davidson BN, et al. A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. EGEMS (Washington, DC). 2016;4(1):1244.

21. Tcheng JE. Optimizing Strategies for Clinical Decision Support: Summary of a Meeting Series: National Academy of Medicine; 2018.

22. Jean-Jacques M, Persell SD, Thompson JA, Hasnain-Wynia R, Baker DW. Changes in disparities following the implementation of a health information technology-supported quality improvement initiative. Journal of general internal medicine. 2012;27(1):71-7.

23. Cholan RA, Weiskopf NG, Rhoton DL, Colin NV, Ross RL, Marzullo MN, et al. Specifications of Clinical Quality Measures and Value Set Vocabularies Shift Over Time: A Study of Change through Implementation Differences. AMIA Annual Symposium proceedings AMIA Symposium. 2017;2017:575-84.