NCQA Relative Resource Utilization Data Architecture Redesign

General Principles for the Data Architecture Redesign

Our general redesign principles were:

- *Turn horizontal to vertical (cube it):* Column combinations create unique points vs. esoteric variable names
- Design a self-documenting ID system via defined enumerations
- Enable easy validation and calculation (i.e., easily sum, aggregate based on columns vs. piecing together hundreds of individual variable names)
- Use industry standards to enable use with common tools
- Reduce complexity of structure and operations
- Remove unnecessary data points
  - Calculate desired data points if and when they are needed
- Require machine-to-machine exchanges
  - No more manual (fat-finger) entry
- Require audit to be more than "visual" inspection via User Interface (UI) reviews

The pages that follow walk you through our new design, the use of XML technology and the benefits of this change. To reiterate, we made these changes with *you* in mind!

Tackling the Data Explosion

New Diabetes Measure

RDI Measure with HCC - Data Field Analysis

Data Table			
Sex Groups	Ages Groups	Risk Groups	Total
3	5	13	195

Cells Per Table	Tables	Total Cells
195	55	10725

User Submitted Data (the core data)				
Sex Groups	Ages Groups	Metric Catagories	Risk Groups	Total
2	4	22	13	2288

Calculated Fields			
Type	Cells	Tables	Total
Procs / 1000	195	7	1365
Disc / 1000	195	11	2145
PMPM	195	6	1170
Totals	91	22	2002
Total Totals	195	9	1755
			8437

Building Blocks

Sex Groups	Age Groups	Risk Groups	Metric Category
Male	18-44	1	CADDiagCount
Female	45-54	2	MemberMonMed
Total	55-64	3	MemberMonPhar
	65-75	4	InpatFacCost
	Total	5	InpatEvalAndMgmtCost
		6	OutpatEvalAndMgmtCost
		7	InpatSurgAndProcCost
		8	OutpatSurgAndProcCost
		9	PharmacyCost
		10	EDVisitCount
		11	InpatFacAcuMedDisCount
		12	InpatFacAcuMedDaysCount
		13	InpatFacAcuSurDisCount
			InpatFacAcuSurDaysCount
			InpatFacNonAcuDisCount
			InpatFacNonAcuDaysCount
			CardiacCathCount
			CABGCount
			CarotidEndarCount
			CarotidArtStenDiagCount
			CardiacCompTomoCount
			PCICount

As you can see in this illustration, in the *Relative Resource Use for People With Diabetes (RDI)* measure, the number of data points increased from 722 to over 10,000.

Data points are intuitively built using a defined set of “building blocks.” Each measure has a set of global meta-data data points (eligible population) and a larger data set consisting of combinations of gender group (male, female), age group (18–44), risk group (1–13) and metric category (Pharmacy Cost or PCI Count).

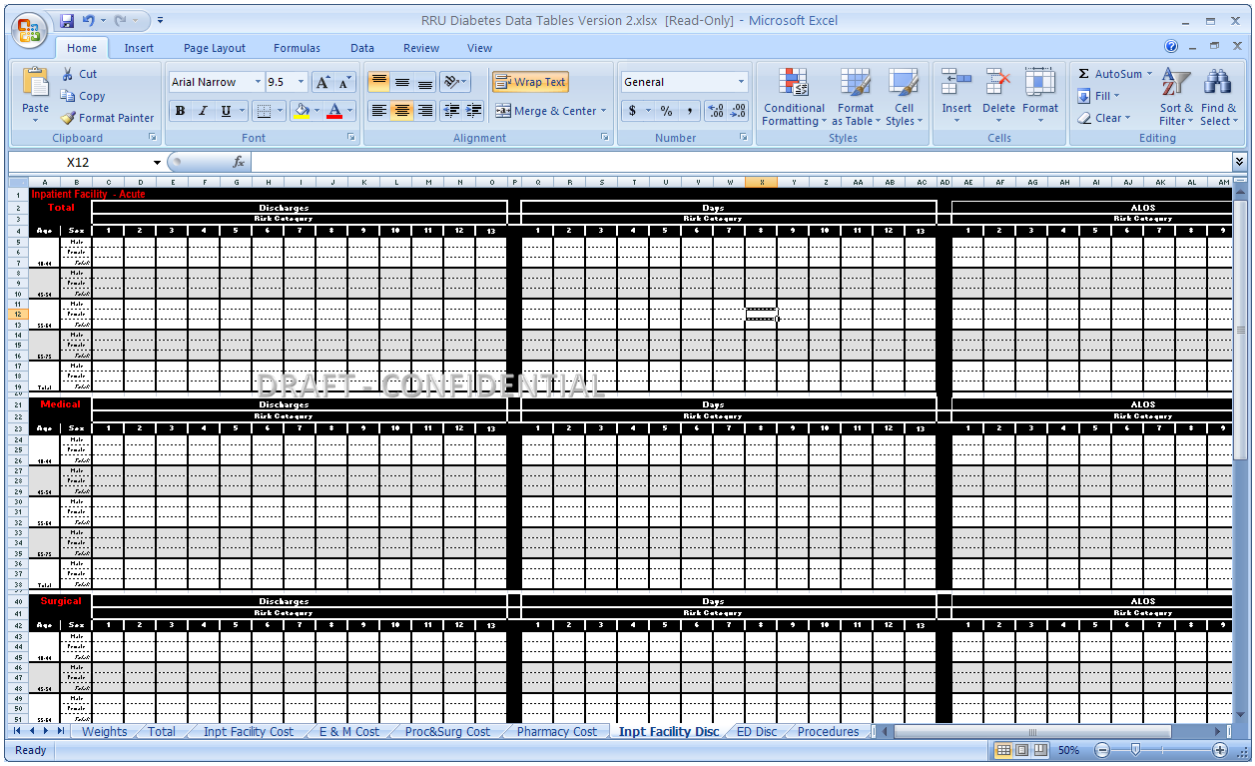
Using these building blocks, you can create an easily identifiable data set of the core “user submitted” data. The calculated fields will no longer need to be defined—they can be created dynamically by user and by need (i.e., if you need to view total-totals, simply aggregate and sum the totals for the stratification you want to see).

Core user-submitted data points break down to four columns of data, each consisting of a clearly defined enumeration of options, plus the value.

Metric Category	Age Group	Gender Group	Risk Category	Value
PharmacyCost	18–44	M	1	NR
PharmacyCost	18–44	M	2	NR
PharmacyCost	18–44	M	3	NR
PharmacyCost	18–44	M	4	NR
PharmacyCost	18–44	M	…	NR
…	…	…	…	NR
EDVisitCount	55–64	F	8	NR
EDVisitCount	55–64	F	9	NR
EDVisitCount	55–64	F	10	NR
EDVisitCount	55–64	F	…	NR

No More Excel-Based Visuals

Because of the increased number of data points, NCQA will no longer provide UI presentations of the data. Visually “viewing” data sets this large is infeasible. The number of screens and cells required for an accurate representation of RDI would require more than 10 Excel sheets, with numerous tables per sheet.



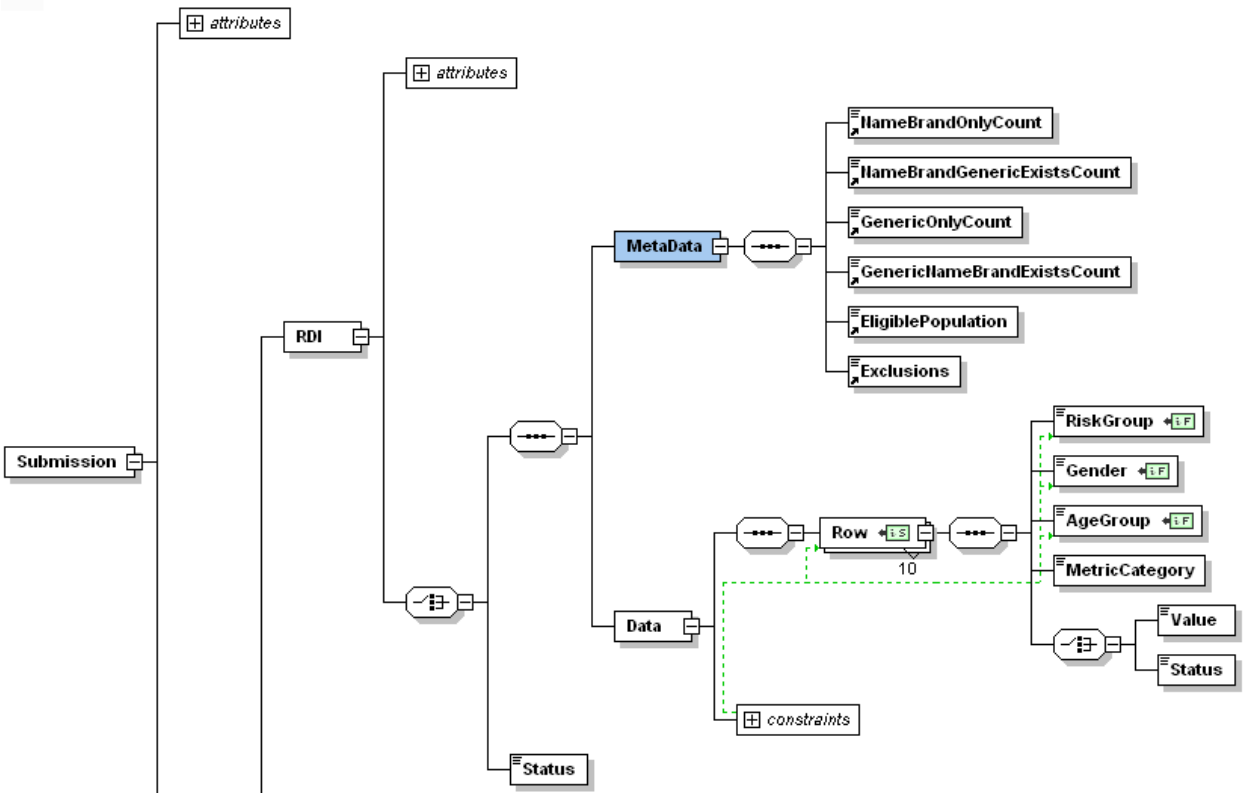
XML Concepts

Although RRU data will still be collected via the Interactive Data Submission System (IDSS) using XML, the XML design will change to reflect the new data architecture. Creation, importation and operations with the XML data will be much easier than the previous IDSS XML design for RRU. Before we continue, let us review a few basic XML concepts.

- *Extensible Markup Language (XML)*. Profile of a Standard Generalized Mark Language (SGML) (1998). A set of rules to create machine readable data. http://www.w3.org/XML/
- *XML Schema Document (XSD)*. Schema, a set of rules to define, constrain and validate XML (2001) http://www.w3.org/XML/Schema
- *Extensible Stylesheet Language (XSL)*. Transformation specification to apply "logic" or transform XML (1999) http://www.w3.org/Style/XSL/

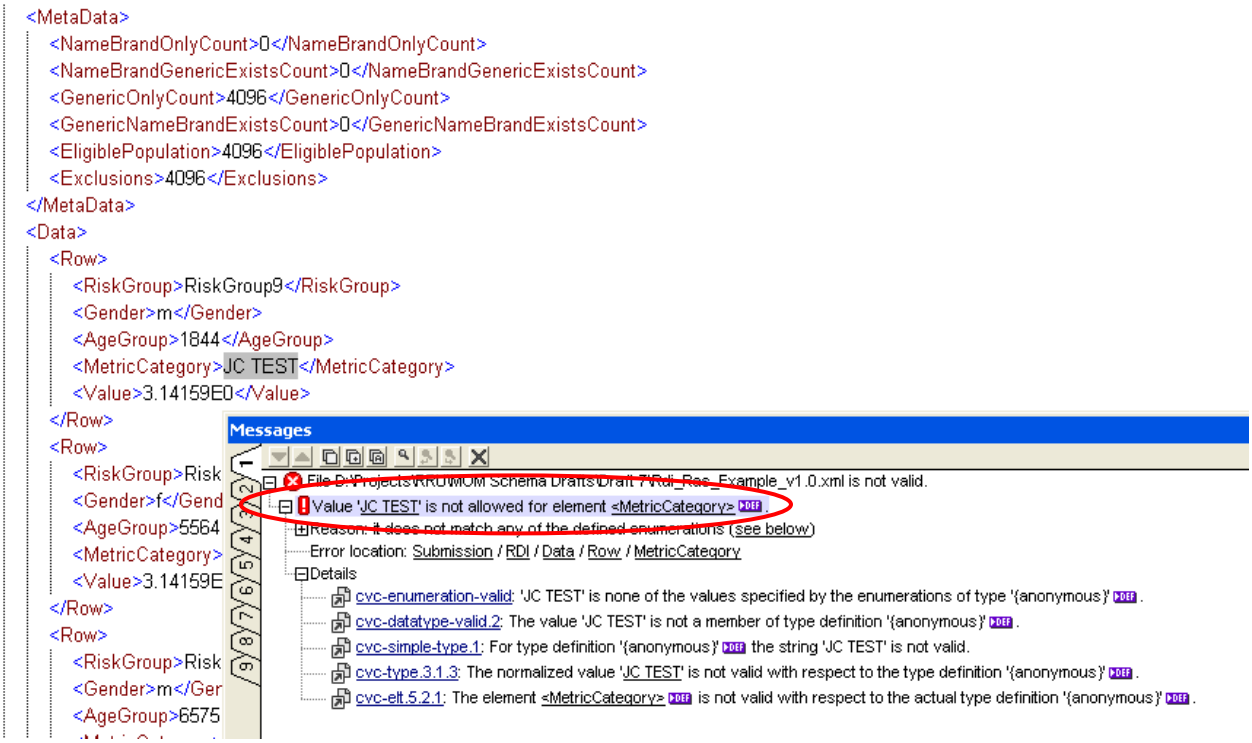
All are widely used within the industry, and many tools support these file types. Even Web browsers have XML engines. For tutorial information, visit: http://www.w3schools.com/xml/xml_whatis.asp.

The New Schema



The new XML schema for collecting RRU data is similar to the building blocks shown above. There is a global node set of meta-data and then a repeating set of “rows” that consist of combinations of building blocks: RiskGroup, Gender, AgeGroup and MetricCategory.

This figure shows a sample XML document that was built using the new XML schema (XSD) file. New to this schema are stronger constraints and defined enumerations per node. Here, the schema was applied to the XML document being viewed with an XML tool (XML Spy). The schema instantly provided a “hint” list of permitted values within the MetricCategory node. Each of the four primary nodes has defined lists that will enable instant data file validation of both structure and content (the options assigned to the nodes).



If an incorrect option is entered into a node, an error message is generated when the XML document is validated (either within a visual tool, as shown here, or by machine validation, as the IDSS uses). The error message in our example is **Value '<u>JC TEST</u>' is not allowed for element <u><MetricCategory></u>**. This type of validation can be (and should be) performed outside of NCQA systems. We will distribute the schema with the Volume 2 Technical Specifications, so that software developers can both configure their systems to the schema and use the schema to help validate the data.

Additional logic can be applied using proprietary coding (as many developers have done in the past) or by simply using another XML technology: *Extensible Stylesheet Language (XSL).* XSL enables you to transform XML into other forms (csv or html reports) and apply additional layers of logic to the XML document. In the example below, we created a validation to check whether ED Visit Counts were "whole numbers." As you can see, they are not, and the XSL generated an error message. Like the XSD (schema) file, these XSL files are technology agnostics. The only thing that is required is an XML parser, which can be found in most modern programming languages, XML tools and even in Explorer and Firefox Web browsers.

```
1  <?xml version="1.0" encoding="UTF-8"?>
2  <Report xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:noNamespaceSchemaLocation="
   C:/Users/maskin.MOMCORP/Desktop/xml31-201/ErrorHandling.xsd">
3    <Errors>
4      <Error Type="Error" Context="RDI,RiskGroup10,m,4554,EDVisitCount,3.14159E0">
5        <message>Counts must be a whole number</message>
6      </Error>
7    </Errors>
8  </Report>
```

Operational Benefits

We know the new data architecture and use of modern XML technology have many benefits for you.

- *XML:* Easier to generate, analyze and work with
- *XSD:* Enforces valid XML and ensures valid values and uniqueness
- *XSL:* Applies additional logic or calculations to XML post XSD validation
- All files can be distributed with RRU specifications, so any party can generate, validate or calculate RRU aggregate data formats external to NCQA, with minimal tools
- Easier, faster collection of data production or pilots (no need to define thousands of variable names)
- NCQA can instantly provide tools to external parties to validate and calculate data more easily
- You only need the basic industry standard tools (XML parser and some knowledge of XML concepts)
- Feeds into NCQA's new "cube" data warehouse design more quickly and easily (faster and reduced risk for seasonal data processing)
- Enables NCQA (and external parties) to change measures more easily (adding one service category is as simple as adding one item to the Metric Category constraint in the schema)
- Less time, less cost, less risk of error for any party working with RRU data