Comments on the caBIG Compatibility document – John Speakman, MSKCC

I have focused on the matrix, in particular the first three categories.

Interface Integration:
- What does "interface integration" mean? Integration with what? Why not just say "interfaces"? Is there some meaning in here that I'm missing?
- In "Legacy" level, what makes a data file "local" and its format "custom"? I guess if its format is not specified in "Bronze", "Silver" or "Gold" then it's "custom" by definition (right?) but what makes it "local"? For instance, I'm assuming that "Legacy" level implies that you can, for instance, extract an ASCII text file and, with the help of its definition, parse it. Or would the fact that a definition exists at all make it "Bronze"? For instance, what would Theradex's ACES be?
- Am I right in thinking that the footnoted definition of "baseline" in "Bronze" means that if two systems are caBIG compatible at the "Bronze" level, and you want these two systems to share data, duplicate data entry need never happen as long as someone does the mapping legwork? As an aside, the issue is of course a bit more complex than that (e.g., our experiences with Theradex) – it might be theoretically possible to eliminate duplicate data entry but in practice sometimes it's simply not worth the effort, so it goes on. It has to be (in general at least) *easier* to implement the interface than to perform duplicate data entry. I guess that's what "Silver" is.
- Mightn't any or all data items need to be shared "to eliminate duplicate data entry and manual processing"? In other words, isn't all data "baseline" data, or at least potentially so?
- In "Silver", what makes an API "well-defined", i.e., how well is well? What's the precise meaning of "access to data objects"?
- What's a "standards-based electronic source"? In other words, we need a list of them.
- Again we need a list of "standard or commonly used interchange formats".
- How well "documented" does the description of the underlying data structures have to be for it to pass muster as "Silver"? What does "data structures that are accessible" mean in this context?
- What's a "standard messaging system"? I know HL7 is one but are there others? When is it "appropriate" – is it any time that the messaging system has a standard for this type of data, i.e., by "appropriate" they mean "applicable"?

Vocabularies / Terminologies and Ontologies:
- Need a list of "publicly accessible standardized controlled vocabularies".
- Is the above (for "Bronze") the same as the "standard terminologies approved by public standards bodies or the caBIG Vocabulary/CDE Workspace" referred to in "Silver"? In other words, are we saying the difference between "Bronze" and "Silver" is just that "Bronze" is "some standard vocabularies" and "Silver" is "100% standard vocabularies"?

- If the above is true, whither "Gold"? Maybe TBD.
- So what about systems that are fully coded (i.e., no free text) but don't use any standard vocabularies? Are they Legacy?

Data Elements:
- How would you "use" metadata for data collection and external reporting? What makes a repository "external" – just that it is "shared by multiple applications"? Is the clinical trials database one application, or multiple applications? If just one big application, obviously it wouldn't be "silver" in terms of interfaces, but the metadata repository wouldn't need to be "external" to be "bronze" in terms of data elements, right?
- Define "controlled terminology" (in "Bronze"). Is it the same as "standard controlled terminologies approved by public standards bodies or the caBIG Vocabulary/CDE Workspace" (in "Silver")? Has the latter approval taken place or must any application seeking "Silver" status wait for it?
- Need a list of standards "comparable" to ISO 11179.
- Again the suggestion that "CDEs are harmonized and re-used from across the Domain Workspace" implies that these CDEs are ready for use in the Domain Workspace, or else that again any application seeking "Silver" Status must wait.

General:

- Following our discussion in San Francisco, the caption to each of Figures 1, 2 and 3 should probably substitute "is an example of" for "illustrates".
- Likewise following that discussion, something needs to be done to make Appendix J more legible.