

**Online Supplemental Table 1.** Framework as at March 2020

Category	Dimension	Definition
<b>Description</b>	Metadata Completeness	Level of metadata completed
	Metadata Quality	Richness of metadata completion – including within required formats and quality of qualitative fields
<b>Scale</b>	Coverage	Number of individuals included in the dataset
	Duration	Length of time to which the data relates
	Depth	Number of information available per individual (e.g. number of fields)
<b>Service</b>	Format	The presentation of the data – to be presented in useful formats / interoperable standards
	Timeliness	How quickly the data can be provided – in a useful timescale
<b>Quality</b>	Completeness	The proportion of data entries that should be populated are populated (and inverse – proportion that should not be populated are not)
	Consistency / Uniformity	Data are presented in the required format and a similar way – e.g. field types, date formats
	Uniqueness	Lack of duplication
	Validity	Data are valid based on acceptable “rules” e.g. age between 0 and 120, pregnancy in male patients, physiological readings within normal ranges
	Accuracy / Verification	The extent to which the data reflects the “real-world”, e.g. level of certainty that fields are accurate
	“Usefulness”	Qualitative, subjective measure by user (e.g. NPS / star rating)
<b>Added Value</b>	Linkage / Mapping	Ability to link with other datasets
	Annotation	Additional fields added to provide further information, including phenotyping