

Online Supplemental Table 2. Interview questions regarding the initial proposed framework

Category	Question
Main focus of the discussion	Can you tell us about an example of a particularly useful or high-quality dataset?
	How did you make use of this dataset?
	Why do you consider that it had these attributes?
	Alternatively, could you recall an experience with an un-useful or low-quality dataset; and the reasons you considered it to be of this nature?
Other points to discuss	As a data user, how do you imagine that a data utility framework/metrics/scores might be able to serve you, support you to make better use of datasets on the innovation gateway?
	Can you describe to us how you might imagine this information working?
	Based on your review of the framework, are there any dimensions that stand out to you as useful/not useful? Why/why not?
	Are there any dimensions that you have questions about, or aren't self-explanatory?
	For your own data needs and use cases, which dimensions would you consider most important/least important?
	For each dimension on the list, how would you rate their importance in terms of understanding data quality
	Are there any other utility or quality dimensions that you would add to this list?
	Data Users in your network: We are extending our engagements, and keen to speak with data users to get their feedback and ideas on data utility. Do you have 3-4 data users in your network you could put us in touch with to interview or survey?
Specific clarification points	Format: A particular element that HDR UK are keen to understand is whether your organisation would be able to comply with a requirement to provide data according to either a standard model and format (e.g. OMOP, or a requirement to make data available through a FHIR API) would the organisation be able to do this now? If not, what would be required for you to be able to do this? Is this already on organisational roadmaps? Is it something that could be feasible within a year or two? Or impossible without significant additional investment?
	Coverage: The suggestion here is "Number of individuals included in the dataset". What would be your requirements in terms of quickly understanding coverage (e.g. number of observations, sites etc)?
	Usefulness: In what format would subjective user feedback on the dataset be useful to you? Reviews? Five-star ratings?

	<p>Validation or Transformation: Level of manual “cleaning” and Annotation: Additional fields added to provide further information, including phenotyping: General feedback on data quality often features statements such as the above, which doesn’t specify the outcome. What specific indicators would be useful to you in relation to these statements?</p>
--	---