# What the Instrument Already Knows: Surfacing Blaise Metadata for Documentation and Discovery


## Abstract

Survey documentation is usually produced after data collection ends, by hand, from sources that have already drifted apart. The Blaise data model records exactly what was asked, in what order, under what conditions, and in which languages, yet that knowledge is hard to reach once the instrument is in production. Codebooks get re-typed from specification documents, question wording is paraphrased, routing context is lost, and the resulting catalog cannot answer the question researchers actually ask: which question produced this variable?

This paper argues that documentation should be a byproduct of collection rather than a separate deliverable produced afterwards. We describe a production pipeline in which questionnaire metadata is captured once as DDI Lifecycle, published to Blaise 5 for data collection, and deposited in the same act into Colectica Repository for publication. Because instrument and documentation derive from one versioned source, every variable in the published catalog carries a resolvable link to its source question, the response options as administered, the universe, the concept measured, and the routing context in which the question appeared.

We then examine what this makes possible downstream. The repository web portal supports variable-level search across studies, question-in-context display, concordance views aligning comparable variables across waves and surveys, and full version history for every item. The same metadata is exposed for reuse through a REST API, OAI-PMH in DDI Lifecycle 3.3 and Dublin Core, and RDF, so holdings are harvestable by external discovery services rather than confined to a single website.

We report practical lessons from production deployments, including how to treat instruments already written in Blaise, where automated extraction suffices, and where human curation remains necessary. We close on governance and transparency, including the value of machine-actionable question metadata as trustworthy grounding for AI-assisted discovery.