# DDI4 Prototype Functionality Definition

Orange: Indicator of not being a show stopper

Red - Priority

## Use Cases

* Data Capture
	+ Simple questionnaire
	+ Simple protocol
	+ Obtaining register data (official data)
	+ Internet source (e.g. web scraping) - requires workflow and protocol for information captured about the web scraping event. Transformation path from source to dataset
	+ Streaming data source (e.g. smart meter data, IOT) - what is limited due to requirement of qualitative
* Publish Common Data Elements
* Describe a dataset
	+ Streaming data source (e.g. smart meter data, IOT) - what is limited due to requirement of qualitative
	+ Describe slices of a dataset
* Describing register data (e.g. RAIRD) - Register Utiliser Tool (Swedish Research Council)
* Statistical Classification
* Data Transformation Processing
	+ Transformation of data from point of capture to resultant dataset
	+ Verification processing during archive ingest
	+ Confidentialization
	+ Imputation
* Describe a simple study
	+ Including discovery and dissemination (ability to populate DISCO)
	+ Ability to create Intentional or *ad hoc* grouping of studies aka 3.2
* Archiving a study (description, provenance, ownership)

### Study Level Information

(should support both 2.5 and 3.2. Current content is focused on Codebook functionality.)

* Citation
* Coverage
* Author and Organisation stuff

### Data Description

* Unit record, multiple records per case
	+ Rectangular, fixed or delimited file
	+ Hierarchical files
* Event Records or Spell data (needs InstanceVariable datatype)
* Aggregate data, mapping to SDMX
* CSVW compatibility - reframe not imported into model directly
	+ Core features, see the CSVW use cases, identify the ones we can do.
	+ Not clear if possible for direction of data columns, list in cell
	+ See Jira DDI4Data-1
* Codebook (Nesstar, CESSDA list)
* Basic descriptives, summary statistics, category statistics
* Variable cascade
* Weights, (instance variables should have a list of references to possible weight variables)
* Streaming data / data sharding / slices

### Data Collection Information

* Mode (Web,PAPI, CAPI etc)
* Coverage (topical, temporal, spatial)

### Data Capture

* Basic instruments with response domains and flow
* Extend description of [https://ddi-alliance.atlassian.net/wiki/spaces/DDI4/pages/491683/Data+Capture+Team](https://ddi-alliance.atlassian.net/wiki/spaces/DDI4/pages/491683/Data%2BCapture%2BTeam)
	+ Survey
	+ Assay
	+ Multi-mode
* To include response domains
* Split / joins

### Discovery

* Can we do full DISCO? Mapping and see gaps

### Data Processing

* Data transformation - describe derived variable, data cleaning, workflow
* Can we support the 3.2 functionality of textual description/computer code etc

### Content Compatibility with previous DDI

* Cross check with 2.5 and 3.2
	+ Archive specific lifecycle information needed? Needs to be done carefully in coordination with creating an execution layer for process
	+ Geography (will be done anyway in the timeframe, relationship to ISO should documented on the item level
* Mapping of DDI4 items to 2.5 and 3.2

## Out of scope for June

* Qualitative
* Complex questionnaires
	+ going off somewhere else
	+ back end office integration
	+ Question Grids
* Execution layer
	+ Elaborate a better description of this not runing a universal computing engine
* Access restrictions on metadata
* Lifecycle orchestration

## Open questions

* Do we need an example for each use case? Does the lack of an example result in removal of the use case from the prototype?
* Assessment of status of existing content
	+ What has not been done?
	+ Approval process - clarify criteria, process, effect of a particular status
	+ Estimate of time to complete each
	+ Prioritization of effort
* Production process and technical outputs (bindings, documentation, etc.)
* Internal review process - content and implementation perspective
* When do we freeze the production process?
* When do we freeze the content? End of March
* When does modeling stop? End of January
* How do test cases fit into all of this - on going validation?

## Timeline:

|  |  |  |  |
| --- | --- | --- | --- |
| Owner | Month Period | Work activities | Deadlines |
| Business Modelers | November | Identified parts of the model that are included in the prototype | End of EDDI: End business modeling in Drupal |
| Modeling Team | December/ January | Review of content, revision for bugs, identification of documentation/example requirements | End of January: Freeze structural content in Drupal |
| Documentation / Processing Team  | February | Work on documentation and binding content and output |  |
| Documentation / Processing Team | March | Work on documentation and binding content and output | NADDI (bgn or end): Freeze binding and documentation |
| TC owns model, Documentation owns examples | April | TC does model integrity review | Examples done by the end April |
| TC owns | May | TC continues model review and reviews documentation and example | Finalize by end of IASSIST |
| TC owns | June | Final preparation for release for review |  |