User's perspective: Review of DDI-Codebook – its current use and future requirements

Knutholmen Sprint | 23–27 May 2016

Goal for the week:

* Identify codebook items used by different institutions
* Identify codebook items required for base codebook view

Next steps:

Follow up:

Links:

# Basic codebook items used by different institutions

## Stanford University:

*Unique perspective:* Faculty view of what is needed to make their lives easier

*Based on:* ICPSR data deposit form

*Unique objects:* Errata, notes of items of interest, Bibliography, News Coverage

## Inter-university Consortium for Political and Social Research (ICPSR):

*Unique perspective:* Captures information from both social science researchers but also topical archives (and the requirements defined by those funders).

*Based on:* ICPSR data deposit form (DDI-C Based)

*Unique objects:* Searchable Database of Archived Data-Related Literature

## SND (Swedish National Data Services):

*Unique perspective:* European perspective (CESSDA); Metadata management; gathering information from the depositor, not necessarily the PI

*Based on:* new deposit form (Use 2.5 and 3.2)

*Unique objects:* Access level

## DNA (Danish National Archive):

*Unique perspective:* National level; researchers in universities are required to deposit; working with National Statistical Organizations

*Based on:* Rectangular data (separating data bases into tables) & qualitative data

*Unique objects: Searchable Database of Research Data*

## IHSN (International Household Survey Network):

*Unique perspective:* Nesstar codebook modified to support NSO work; unique client base and role on promoting standards, complete documentation, and data sharing

*Based on:* DDI 2.5 (many of the addition elements were requested by IHSN)

*Unique objects:* Official title and author of survey; authorization, dates of collection period, subjects, coverage geography (other that countries); units of analysis, sampling, collection mode

# Functional View documentation

## Purpose:

A basic set of objects that can support structured searching for discovery and ingest to statistical software, and provide sufficient information for the intelligent use and understanding of the associated data. Making data independently understandable so people can reuse it.

## Use Cases:

This Functional View supports three major uses

1. Basic search to find data (general search, known item search). Users are looking to use data.
2. Machine actionable to allow users to analyse the data (ingest to statistical software)
3. Informational content to support Intelligent use and understanding the data

**Target Audience:**

Individual researchers, research groups, National Statistical Organizations focusing on stand-alone codebooks, students, instructors, archives, and libraries.

## 

**Table of Objects Selected for Basic Codebook View Based on Three Uses: General Search, Understand the Data, and Machine Actionable (1=needed; 0=not needed)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Object** | **Descriptive Path** | **General Search** | **Understand the Data** | **Machine Actionable** |
| **Document Description** |  |  |  |  |
| Title | DocumentDescription\_Citation\_TitleStatement\_Titl | 1 | 1 | 1 |
| IDNumber | DocumentDescription\_Citation\_TitleStatement\_IDNumbe | 1 | 1 | 1 |
| Authors | DocumentDescription\_Citation\_Authors | 1 | 1 | 1 |
| Producers | DocumentDescription\_Citation\_ProductionStatement\_Produc | 1 | 1 | 1 |
| Copyright | DocumentDescription\_Citation\_ProductionStatement\_Copyri | 1 | 1 | 1 |
| Contacts | DocumentDescription\_Citation\_DistributionStatement\_Contact | 1 | 1 | 1 |
| ProductionDate | DocumentDescription\_Citation\_ProductionStatement\_Produc | 1 | 1 | 1 |
| Text | DocumentDescription\_Citation\_Version\_Tex | 1 | 1 | 1 |
| **Study Description** |  |  |  |  |
| Title | StudyDescription\_Citation\_TitleStatement\_Titl | 1 | 1 | 1 |
| SubTitle | StudyDescription\_Citation\_TitleStatement\_SubTitl | 1 | 1 | 1 |
| AlternativeTitle | StudyDescription\_Citation\_TitleStatement\_AlternativeTit | 1 | 1 | 1 |
| ParallelTitle | StudyDescription\_Citation\_TitleStatement\_ParallelTitl | 1 | 1 | 1 |
| IDNumber | StudyDescription\_Citation\_TitleStatement\_IDNumbe | 1 | 1 | 1 |
| Authors with affiliation | StudyDescription\_Citation\_Authors | 1 | 1 | 1 |
| Acknowledgments | StudyDescription\_Citation\_Acknowledgment | 0 | 1 | 0 |
| Producers with abbr and affiliation (see next) | StudyDescription\_Citation\_ProductionStatement\_Producer | 1 | 1 | 1 |
| Producer affiliation | StudyDescription\_Citation\_ProductionStatement\_Producer\_affiliation | 1 | 1 | 1 |
| Copyright | StudyDescription\_Citation\_ProductionStatement\_Copyrigh | 1 | 1 | 1 |
| ProductionDate | StudyDescription\_Citation\_ProductionStatement\_Production | 0 | 1 | 1 |
| FundingAgency | StudyDescription\_Citation\_ProductionStatement\_FundingAgency | 1 | 1 | 1 |
| Grant Number | StudyDescription\_Citation\_ProductionStatement\_GrantNumber | 1 | 1 | 1 |
| Distributors | StudyDescription\_Citation\_DistributorStatement\_Distributor | 1 | 1 | 1 |
| SeriesName | StudyDescription\_Citation\_SeriesName | 1 | 1 | 1 |
| SeriesInformation | StudyDescription\_Citation\_SeriesInformatio | 0 | 1 | 1 |
| Text | StudyDescription\_Citation\_Version\_Tex | 0 | 1 | 1 |
| Date | StudyDescription\_Citation\_Version\_Date | 1 | 1 | 1 |
| BiblCitation | StudyDescription\_Citation\_BiblCitation | 0 | 0 | 0 |
| Keywords | StudyDescription\_Scope\_Keywords | 1 | 1 | 1 |
| TopicClassifications | StudyDescription\_Scope\_TopicClassifications | 1 | 1 | 1 |
| Abstract | StudyDescription\_Scope\_Abstrac | 1 | 1 | 1 |
| TimePeriod | StudyDescription\_Scope\_SummaryDataDescription\_TimePe | 1 | 1 | 1 |
| CollectionDates | StudyDescription\_Scope\_SummaryDataDescription\_Collectio | 1 | 1 | 1 |
| Countries | StudyDescription\_Scope\_SummaryDataDescription\_Countrie | 1 | 1 | 1 |
| GeographicCoverage | StudyDescription\_Scope\_SummaryDataDescription\_Geogra | 1 | 1 | 1 |
| GeographicUnit | StudyDescription\_Scope\_SummaryDataDescription\_Geogra | 1 | 1 | 1 |
| GeographicBoundingBox | StudyDescription\_Scope\_SummaryDataDescription\_Geogra | 1 | 1 | 1 |
| GeographicPolygon | StudyDescription\_Scope\_SummaryDataDescription\_Geogra | 0 | 1 | 1 |
| AnalysisUnit | StudyDescription\_Scope\_SummaryDataDescription\_Analysis | 1 | 1 | 1 |
| Universe | StudyDescription\_Scope\_SummaryDataDescription\_Univers | 1 | 1 | 1 |
| DataKind | StudyDescription\_Scope\_SummaryDataDescription\_DataKin | 1 | 1 | 1 |
| Notes | StudyDescription\_Scope\_Notes | 0 | 0 | 0 |
| TimeMethod | StudyDescription\_Methodology\_TimeMetho | 1 | 1 | 1 |
| DataCollectors | StudyDescription\_Methodology\_DataCollector | 0 | 1 | 1 |
| Frequency | StudyDescription\_Methodology\_Frequenc | 1 | 1 | 1 |
| Sampling | StudyDescription\_Methodology\_Samplin | 1 | 1 | 1 |
| CollectionMethod | StudyDescription\_Methodology\_CollectionMetho | 1 | 1 | 1 |
| ResearchInstrument | StudyDescription\_Methodology\_ResearchInstrumen | 1 | 1 | 1 |
| CollectionSituation | StudyDescription\_Methodology\_CollectionSituatio | 0 | 1 | 1 |
| Weighting | StudyDescription\_Methodology\_Weighting | 0 | 1 | 1 |
| CleaningOperations | StudyDescription\_Methodology\_CleaningOperation | 0 | 1 | 1 |
| Notes | StudyDescription\_Methodology\_Note | 0 | 0 | 0 |
| ResponseRate | StudyDescription\_Methodology\_ResponseRat | 0 | 1 | 1 |
| StorageLocations | StudyDescription\_AccessConditions\_Availability\_StorageLoc | 0 | 0 | 1 |
| Status | StudyDescription\_AccessConditions\_Availability\_Statu | 1 | 1 | 1 |
| ConfidentialityDeclaration\_Tex | StudyDescription\_AccessConditions\_UseStatement\_Confide | 0 | 1 | 1 |
| Restrictions | StudyDescription\_AccessConditions\_UseStatement\_Restrict | 0 | 1 | 1 |
| Contacts | StudyDescription\_AccessConditions\_UseStatement\_Contact | 0 | 1 | 1 |
| CitationRequiremen | StudyDescription\_AccessConditions\_UseStatement\_Citation | 0 | 1 | 1 |
| DepositRequiremen | StudyDescription\_AccessConditions\_UseStatement\_Deposit | 0 | 1 | 1 |
| Conditions | StudyDescription\_AccessConditions\_UseStatement\_Conditio | 1 | 1 | 1 |
| Disclaimer | StudyDescription\_AccessConditions\_UseStatement\_Disclaim | 0 | 1 | 0 |
| **File Description** |  |  |  |  |
| MissingData | FileDescription\_MissingData | 0 | 1 | 1 |
| Notes | FileDescription\_Notes | 0 | 0 | 0 |
| **Variable Description** |  |  |  |  |
| VariableGroup\_VarList | DataDscr\_varGrp\_var | 1 | 1 | 1 |
| VariableGroup\_VarGroups | DataDscr\_varGrp\_varGrp | 1 | 1 | 1 |
| VariableGroup\_Label | DataDscr\_varGrp\_Labl | 1 | 1 | 1 |
| VariableName | Var\_name | 1 | 1 | 1 |
| VariableFiles | Var\_files | 1 | 1 | 1 |
| VariableIntrvl | var\_intrvl | 0 | 1 | 1 |
| VariableLocationStart | var\_location\_StartPos | 0 | 1 | 1 |
| VariableLocationEnd | var\_location\_EndPos | 0 | 1 | 1 |
| VariableLocationWidth | var\_location\_width | 0 | 1 | 1 |
| VariableLocationrecSegNo | var\_location\_recSegNo | 0 | 1 | 1 |
| VariableLabel | var\_labl | 1 | 1 | 1 |
| Imputation | VarDoc\_Imputation | 1 | 1 | 1 |
| Security | VarDoc\_Security | 0 | 1 | 1 |
| Embargo | VarDoc\_Embargo | 0 | 1 | 1 |
| EmbargoDate | VarDoc\_EmbargoDate | 0 | 1 | 1 |
| ResponseUnit | VarDoc\_ResponseUnit | 0 | 1 | 1 |
| AnalysisUnit | VarDoc\_AnalysisUnit | 0 | 1 | 1 |
| PreQuestion | VarDoc\_PreQuestion | 0 | 1 | 1 |
| LiteralQuestion | VarDoc\_LiteralQuestion | 1 | 1 | 1 |
| PostQuestion | VarDoc\_PostQuestion | 0 | 1 | 1 |
| InterviewerInstructions | VarDoc\_InterviewerInstructions | 0 | 1 | 1 |
| SummaryStatistics | var\_sumStat | 0 | 1 | 1 |
| Text | VarDoc\_Text | 0 | 0 | 0 |
| Var/Notes | At the variable level | 0 | 1 | 1 |
| DerivationDescription | var\_derivation | 0 | 1 | 1 |
| Var/DrvDsc | Variable derivation description | 0 | 1 | 1 |
| Var/DrvComd | Variable derivation commands | 0 | 1 | 1 |
| Variable concept |  | 1 | 1 | 1 |
| Variable concept description |  | 1 | 1 | 1 |
| Var - measurement unit (Attribute) |  | 1 | 1 | 1 |
| Var/VarCat | variable categories | 1 | 1 | 1 |
| Var/catgry | Value labels | 1 | 1 | 1 |
| Question flow |  | 0 | 1 | 1 |
|  |  |  |  |  |
| Total Needed |  | 54 | 90 | 89 |

After smaller group discussions and reviews by the Technical groups, we revised our use cases to reflect the elements required for a Basic Search, an Advanced Search, and elements required for a deeper Understanding of the metadata and data. These are reflected in the following table:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Object** | **Descriptive Path** | **Basic Search** | **Advanced Search** | **Under-**  **standing** |
| **Document Description** |  |  |  |  |
| Title | DocumentDescription\_Citation\_TitleStatement\_Titl | 0 | 0 | 1 |
| IDNumber | DocumentDescription\_Citation\_TitleStatement\_IDNumbe | 0 | 0 | 1 |
| Authors | DocumentDescription\_Citation\_Authors | 0 | 0 | 1 |
| Producers | DocumentDescription\_Citation\_ProductionStatement\_Produc | 0 | 0 | 1 |
| Copyright | DocumentDescription\_Citation\_ProductionStatement\_Copyri | 0 | 0 | 1 |
| Contacts | DocumentDescription\_Citation\_DistributionStatement\_Contact | 0 | 0 | 1 |
| ProductionDate | DocumentDescription\_Citation\_ProductionStatement\_Produc | 0 | 0 | 1 |
| Text | DocumentDescription\_Citation\_Version\_Tex | 0 | 0 | 1 |
| **Study Description** |  |  |  |  |
| Title | StudyDescription\_Citation\_TitleStatement\_Titl | 1 | 1 | 1 |
| SubTitle | StudyDescription\_Citation\_TitleStatement\_SubTitl | 1 | 1 | 1 |
| AlternativeTitle | StudyDescription\_Citation\_TitleStatement\_AlternativeTit | 1 | 1 | 1 |
| ParallelTitle | StudyDescription\_Citation\_TitleStatement\_ParallelTitl | 1 | 1 | 1 |
| IDNumber | StudyDescription\_Citation\_TitleStatement\_IDNumbe | ? | 1 | 1 |
| Authors with affiliation | StudyDescription\_Citation\_Authors | 1 | 1 | 1 |
| Acknowledgments | StudyDescription\_Citation\_Acknowledgment | 0 | 0 | 0 |
| Producers with abbr and affiliation (see next) | StudyDescription\_Citation\_ProductionStatement\_Producer | 1 | 1 | 1 |
| Producer affiliation | StudyDescription\_Citation\_ProductionStatement\_Producer\_affiliation | 1 | 1 | 1 |
| Copyright | StudyDescription\_Citation\_ProductionStatement\_Copyrigh | 0 | 1 | 1 |
| ProductionDate | StudyDescription\_Citation\_ProductionStatement\_Production | 0 | 0 | 1 |
| FundingAgency | StudyDescription\_Citation\_ProductionStatement\_FundingAgency | 0 | 1 | 1 |
| Grant Number | StudyDescription\_Citation\_ProductionStatement\_GrantNumber | 0 | 1 | 1 |
| Distributors | StudyDescription\_Citation\_DistributorStatement\_Distributor | 0 | 1 | 1 |
| SeriesName | StudyDescription\_Citation\_SeriesName | 1 | 1 | 1 |
| SeriesInformation | StudyDescription\_Citation\_SeriesInformatio | 0 | 0 | 1 |
| Text | StudyDescription\_Citation\_Version\_Tex | 0 | 0 | 1 |
| Date | StudyDescription\_Citation\_Version\_Date | 0 | 1 | 1 |
| BiblCitation | StudyDescription\_Citation\_BiblCitation | 0 | 0 | 0 |
| Keywords | StudyDescription\_Scope\_Keywords | 1 | 1 | 1 |
| TopicClassifications | StudyDescription\_Scope\_TopicClassifications | 1 | 1 | 1 |
| Abstract | StudyDescription\_Scope\_Abstrac | 1 | 1 | 1 |
| TimePeriod | StudyDescription\_Scope\_SummaryDataDescription\_TimePe | 1 | 1 | 1 |
| CollectionDates | StudyDescription\_Scope\_SummaryDataDescription\_Collectio | 0 | 1 | 1 |
| Countries | StudyDescription\_Scope\_SummaryDataDescription\_Countrie | 1 | 1 | 1 |
| GeographicCoverage | StudyDescription\_Scope\_SummaryDataDescription\_Geogra | 1 | 1 | 1 |
| GeographicUnit | StudyDescription\_Scope\_SummaryDataDescription\_Geogra | 1 | 1 | 1 |
| GeographicBoundingBox | StudyDescription\_Scope\_SummaryDataDescription\_Geogra | 0 | 1 | 1 |
| GeographicPolygon | StudyDescription\_Scope\_SummaryDataDescription\_Geogra | 0 | 0 | 1 |
| AnalysisUnit | StudyDescription\_Scope\_SummaryDataDescription\_Analysis | 1 | 1 | 1 |
| Universe | StudyDescription\_Scope\_SummaryDataDescription\_Univers | 1 | 1 | 1 |
| DataKind | StudyDescription\_Scope\_SummaryDataDescription\_DataKin | 1 | 1 | 1 |
| Notes | StudyDescription\_Scope\_Notes | 0 | 0 | 0 |
| TimeMethod | StudyDescription\_Methodology\_TimeMetho | 0 | 1 | 1 |
| DataCollectors | StudyDescription\_Methodology\_DataCollector | 0 | 0 | 1 |
| Frequency | StudyDescription\_Methodology\_Frequenc | 1 | 1 | 1 |
| Sampling | StudyDescription\_Methodology\_Samplin | 0 | 1 | 1 |
| CollectionMethod | StudyDescription\_Methodology\_CollectionMetho | 0 | 1 | 1 |
| ResearchInstrument | StudyDescription\_Methodology\_ResearchInstrumen | 0 | 1 | 1 |
| CollectionSituation | StudyDescription\_Methodology\_CollectionSituatio | 0 | 0 | 1 |
| Weighting | StudyDescription\_Methodology\_Weighting | 0 | 0 | 1 |
| CleaningOperations | StudyDescription\_Methodology\_CleaningOperation | 0 | 0 | 1 |
| Notes | StudyDescription\_Methodology\_Note | 0 | 0 | 0 |
| ResponseRate | StudyDescription\_Methodology\_ResponseRat | 0 | 0 | 1 |
| StorageLocations | StudyDescription\_AccessConditions\_Availability\_StorageLoc | 0 | 0 | 1 |
| Status | StudyDescription\_AccessConditions\_Availability\_Statu | ? | 1 | 1 |
| ConfidentialityDeclaration\_Tex | StudyDescription\_AccessConditions\_UseStatement\_Confide | 0 | 0 | 1 |
| Restrictions | StudyDescription\_AccessConditions\_UseStatement\_Restrict | 0 | 0 | 1 |
| Contacts | StudyDescription\_AccessConditions\_UseStatement\_Contact | 0 | 0 | 1 |
| CitationRequiremen | StudyDescription\_AccessConditions\_UseStatement\_Citation | 0 | 0 | 1 |
| DepositRequiremen | StudyDescription\_AccessConditions\_UseStatement\_Deposit | 0 | 0 | 1 |
| Conditions | StudyDescription\_AccessConditions\_UseStatement\_Conditio | 0 | 1 | 1 |
| Disclaimer | StudyDescription\_AccessConditions\_UseStatement\_Disclaim | 0 | 0 | 0 |
| **File Description** |  |  |  |  |
| MissingData | FileDescription\_MissingData | 0 | 0 | 1 |
| Notes | FileDescription\_Notes | 0 | 0 | 0 |
| **Variable Description** |  |  |  |  |
| VariableGroup\_VarList | DataDscr\_varGrp\_var | 0 | 1 | 1 |
| VariableGroup\_VarGroups | DataDscr\_varGrp\_varGrp | 0 | 1 | 1 |
| VariableGroup\_Label | DataDscr\_varGrp\_Labl | 0 | 1 | 1 |
| VariableName | Var\_name | 1 | 1 | 1 |
| VariableFiles | Var\_files | 0 | 1 | 1 |
| VariableIntrvl | var\_intrvl | 0 | 0 | 1 |
| VariableLocationStart | var\_location\_StartPos | 0 | 0 | 1 |
| VariableLocationEnd | var\_location\_EndPos | 0 | 0 | 1 |
| VariableLocationWidth | var\_location\_width | 0 | 0 | 1 |
| VariableLocationrecSegNo | var\_location\_recSegNo | 0 | 0 | 1 |
| VariableLabel | var\_labl | 1 | 1 | 1 |
| Imputation | VarDoc\_Imputation | 0 | 1 | 1 |
| Security | VarDoc\_Security | 0 | 0 | 1 |
| Embargo | VarDoc\_Embargo | 0 | 0 | 1 |
| EmbargoDate | VarDoc\_EmbargoDate | 0 | 0 | 1 |
| ResponseUnit | VarDoc\_ResponseUnit | 0 | 0 | 1 |
| AnalysisUnit | VarDoc\_AnalysisUnit | 0 | 0 | 1 |
| PreQuestion | VarDoc\_PreQuestion | 0 | 0 | 1 |
| LiteralQuestion | VarDoc\_LiteralQuestion | 1 | 1 | 1 |
| PostQuestion | VarDoc\_PostQuestion | 0 | 0 | 1 |
| InterviewerInstructions | VarDoc\_InterviewerInstructions | 0 | 0 | 1 |
| SummaryStatistics | var\_sumStat | 0 | 0 | 1 |
| Text | VarDoc\_Text | 0 | 0 | 0 |
| Var/Notes | At the variable level | 0 | 0 | 1 |
| DerivationDescription | var\_derivation | 0 | 0 | 1 |
| Var/DrvDsc | Variable derivation description | 0 | 0 | 1 |
| Var/DrvComd | Variable derivation commands | 0 | 0 | 1 |
| Variable concept |  | 1 | 1 | 1 |
| Variable concept description |  | 0 | 1 | 1 |
| Var - measurement unit (Attribute) |  | 0 | 1 | 1 |
| Var/VarCat | variable categories | 0 | 1 | 1 |
| Var/catgry | Value labels | 1 | 1 | 1 |
| Question flow |  | 0 | 0 | 1 |
|  |  |  |  |  |
| Total Needed |  | 24 | 46 | 89 |