Dimensions & Metrics Integrated into the Conversational Agent for Open Government Data Quality Assessment

Dimensions	Metrics	Definition	Formula
Metadata			
Completeness	Percentage of complete fields in metadata	Indicates the percentage of complete fields in metadata. A complete field is defined as a value that is not null and is set	Number of complete fields * 100/Total number of fields
Findability	Keywords assigned	Indicates if the keywords field is assigned	100 if yes else 0
	Categories assigned	Indicates if the categories field is assigned	100 if yes else 0
	Title given	Indicates if the dataset title is set	100 if yes else 0
	Description of data given	Indicates if the dataset description is given	100 if yes else 0
	Temporal information given	Indicates if the date/period of the dataset belongs to is given	100 if yes else 0
	Spatial information given	Indicates if the information about the region the dataset belongs to is given	100 if yes else 0
Accessibility	Access URL accessible	Indicates if the access URL of the dataset is accessible	100 if yes else 0
	Download URL given	Indicates if the download URL of the dataset is given	100 if yes else 0
	Download URL accessible without registration	Indicates if the download URL of the dataset is accessible without registration	100 if yes else 0
Conformity	Conformity of URLs	The extent to which the values of access properties (HTTP, URLs) are valid	Number of valid URL fields * 100/Total URL fields
	Conformity of date formats	The extent to which the date fields are using a valid date format	Number of valid date fields * 100/Total number of date fields
	Conformity of email addresses	The extent to which the email fields are valid	Number of valid email fields *100/Total number of email fields
	DCAT-AP compliance of metadata	Indicates if the metadata of the dataset is provided in DCAT format and is compliant with DCAT-AP	100 if yes else 0
Machine readability/ processability	Processability of file format and media type	Indicates if the file format and media type can be handled by automated processes (CSV, JSON, Excel accepted in this study)	100 if yes else 0
Openness	Openness of file format and media type	Indicates if the file format and media type is in a non-proprietary format (CSV, JSON accepted in this study)	100 if yes else 0
	License information given	Indicates if the license information is given	100 if yes else 0
	Openness of license	Indicates if the license corresponds to one of the open Licenses listed in opendefinition.org	100 if yes else 0
Timeliness	Update information given	Indicates if the frequency/periodicity of the dataset update is given	100 if yes else 0
	Creation date given	Indicates if the creation date is given	100 if yes else 0
	Modification date given	Indicates if the last modification date is given	100 if yes else 0
Accuracy	File format accuracy	Indicates if information given about file format can be compared with the actual file format of the resource	100 if yes else 0
	Content size accuracy	Indicates if information given about content size can be compared with the actual content size of the resource	100 if yes else 0

Understandability	Percentage of columns with metadata	Indicates the percentage of columns in a dataset that has associated descriptive metadata.	Number of columns with metadata * 100/ Total number of columns
Credibility	Contact point given	Indicates if the contact point of the dataset is given	100 if yes else 0
	Dataset publisher given	Indicates if the dataset publisher information (e.g., name or email) is given	100 if yes else 0
Uniqueness	Title is unique	Indicates if the dataset title is unique	100 if yes else 0
	Description is unique	Indicates if the dataset description is unique	100 if yes else 0
	Identifier is unique	Indicates if the dataset identifier is unique	100 if yes else 0
Data			
Completeness	Percentage of complete cells	A cell which is not empty nor null is considered as complete	Number of complete cells * 100/ Total number of cells
	Percentage of complete rows	A row which all the cells are not empty nor null is considered as complete	Number of complete rows * 100/ Total number of rows
Accuracy	Percentage of accurate cells	Accurate cells are identified by their corresponding column type. Therefore, a cell is deemed accurate if its type matches the column type. For instance, if a column has a type of "email", a cell with the value "abc" in that column would be regarded as inaccurate.	Number of accurate cells * 100/ Total number of cells
Uniqueness	Percentage of unique rows	Indicates the percentage of unique rows in the dataset	Number of unique rows * 100/ Total number of rows
	Percentage of unique columns	Indicates the percentage of unique columns in the dataset	Number of unique columns * 100/ Total number of columns

Table: List of dimensions and metrics integrated into the conversational agent.