ERS
ECONOMIC RESEARCH SERVICE
United States Department of Agriculture

GEORGE MASON UNIVERSITY

THE COMMODITY SUBJECT AREA DATABASE (CoSD)

A group of subject area databases, a public data repository, software tools with advanced interfaces, business rules, data validation and streaming procedures, and data privacy methods for the Market and Trade Economics Division (MTED) at ERS.

THE PRODUCT MANUAL
2016-2017

THE COSD DATABASES

This is the product manual that explains the design and implementation of the **Co**mmodity **S**ubject Area **D**atabases (CoSD); pronounced *kozdy*.

The agricultural commodities are categorized into 10 groups; each group is its own database in our project:

1. Animal Products (AP)
2. Vegetables and Pulses (Veg)
3. Fruit and Tree Nuts
4. Cotton and Fibers
5. Sugar and Sweeteners
6. Oil Crops
7. Wheat (Grains)
8. Rice (Grains)
9. Feed Grain (Grains)
10. Macro Economics (not commodity based)

The CoSDs were implemented with 18 Lookup tables, 3 Aggregation tables, and 6 Data tables.

HISTORICAL DATA MIGRATION

Data from multiple sources is managed: NASS, US Trade, WASDE, BLS, and AMS. The databases will have tools accessing the structured data to help analysts in their research and studies (the tools use cases, workflow and their user interface are already defined and developed in Visual Studio).

To push a table from one database to another, we use the following format:

```
/**Drop the keys linked to the table**/
ALTER TABLE [AnimalProductsCoSD].[CoSD].[ERSDataValues] DROP CONSTRAINT
[FK_ERSDataValues_ERSCollection_ID]


/**Drop the table **/
DROP TABLE [AnimalProductsCoSD].[CoSD].[ERSCollection_LU]


/** push data from source table to destination table **/
SELECT *
INTO [AnimalProductsCoSD].[CoSD].[ERSCollection_LU]
FROM [MasterCoSD].CoSD.[ERSCollection_LU]


/** Set the primary key for the table **/
```

```
ALTER TABLE [AnimalProductsCoSD].[CoSD].[ERSCollection_LU]
ADD CONSTRAINT [PK_ERSCollection_LU] PRIMARY KEY CLUSTERED
(
	[ERSCollection_ID] ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF, ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]


/** Re-enforce the foreign keys on the table **/
ALTER TABLE [AnimalProductsCoSD].[CoSD].[ERSDataValues]  WITH CHECK ADD  CONSTRAINT [FK_ERSDataValues_ERSCollection_ID] FOREIGN KEY([ERSDataValues_ERSCollection_ID])
REFERENCES [AnimalProductsCoSD].[CoSD].[ERSCollection_LU] ([ERSCollection_ID])


ALTER TABLE [AnimalProductsCoSD].[CoSD].[ERSDataValues] CHECK CONSTRAINT [FK_ERSDataValues_ERSCollection_ID]
```

The following query is to be used to migrate Data Values for BLS:

```
SELECT
	 --,[ERSDataValues_ID]
      B.ERSSource_ID AS 'ERSDataValues_ERSSource_ID'
      ,9 AS 'ERSDataValues_ERSOutput_ID' --Other
       ,52 AS 'ERSDataValues_ERSStatisticType_ID1'
      ,'NULL' AS ERSDataValues_ERSStatisticType_ID2
      ,'NULL' AS ERSDataValues_ERSStatisticType_ID3
        , C.ERSTimeDimension_ID  AS 'ERSDataValues_ERSTimeDimension_ID'
     ,G.ERSCommodity_ID AS 'ERSDataValues_ERSCommodity_ID'
      ,5 AS 'ERSDataValues_ERSDataFeedType_ID' -- 5       DB in Building
     ,8 AS 'ERSDataValues_ERSUnit_ID1'
      ,'NULL' AS ERSDataValues_ERSUnit_ID2
      ,'NULL' AS ERSDataValues_ERSUnit_ID3
      ,4746 AS ERSDataValues_ERSGeography_ID
      ,A.Value AS 'ERSDataValues_AttributeValue1'
      ,'NULL' AS ERSDataValues_AttributeValue2
      ,'NULL' AS ERSDataValues_AttributeValue3
    ,4 AS 'ERSDataValues_ERSCollection_ID'
     ,6 AS 'ERSDataRowPrivacy_ID'
      ,[Base_Date]
      ,[Stat_Year]
      ,[Period]
      ,[MonthAbr]
      ,[Display]
```

```
  FROM [StagingISD].[dbo].[BLS_Veg_ProducerIndex] A
  LEFT JOIN VegetablesCoSD.CoSD.ERSSource_LU B ON A.Survey=B.ERSSource_Desc
    INNER JOIN VegetablesCoSD.CoSD.ERSCommodityDataSeries G ON A.Series_Id=
G.ERSCommodity_SourceSeriesID
  LEFT JOIN VegetablesCoSD.CoSD.ERSTimeDimension_LU C
                                    ON A.Stat_Year=C.ERSTimeDimension_Year
                                                AND A.MonthAbr=C.ERSTimeDimension_Month
                                                AND
C.ERSTimeDimension_TimeDimensionType_ID=11
```

To select the time dimension of BLS:

```
SELECT
CASE WHEN (A.stat_year= LEFT(A.Base_Date,4) AND (RIGHT(A.Base_Date,2)= RIGHT(A.Period, 2)OR
(RIGHT(A.Base_Date,2)= 00 AND RIGHT(A.Period, 2) = 13)))
                      THEN (SELECT H1.ERSTimeDimension_ID FROM
VegetablesCoSD.CoSD.ERSTimeDimension_LU H1 WHERE H1.ERSTimeDimension_TimeDimensionType_ID=32)
       ELSE (SELECT H2.ERSTimeDimension_ID FROM VegetablesCoSD.CoSD.ERSTimeDimension_LU H2
WHERE H2.ERSTimeDimension_TimeDimensionType_ID=11 AND A.Stat_Year= H2.ERSTimeDimension_Year
       AND H2.ERSTimeDimension_Month = A.MonthAbr)
END
,A.Base_Date, A.stat_year,H.ERSTimeDimension_Year,A.Period,A.MonthAbr,
H.ERSTimeDimension_Month
FROM [StagingISD].[dbo].[BLS_Veg_ProducerIndex] A
LEFT JOIN VegetablesCoSD.CoSD.ERSTimeDimension_LU H ON A.Stat_Year= H.ERSTimeDimension_Year
       AND H.ERSTimeDimension_Month = A.MonthAbr
WHERE A.[Item_Name] in
              (SELECT distinct  B.[Item_Name] FROM [StagingISD].[dbo].[BLS_Veg_ProducerIndex]
B WHERE B.stat_year= 2015)
```

To insert Data Series for constructed variable:

```
SELECT B.ERSCommoditySubCommodity_ID AS [ERSCommoditySubCommodity_ID]
      ,1 AS [ERSCommodity_ERSSector_ID] -- AP
      ,B.ERSCommoditySubCommodity_GroupID AS [ERSCommodity_ERSGroup_ID]
      ,PhysicalAttribute_ID
      ,PhysicalAttribute_Desc
      ,A.[Stat ID]
      ,ProdPractice_ID
      ,UtilPractice_ID
      ,HS_ID AS [ERSCommodity_ERSHS_ID]
      ,'(' + A.[Data series] + ') ' + A.[Data series IDs] AS [ERSCommodity_SourceSeriesID]
      ,A.Commodity+ ' - ' + A.StatType AS [ERSCommodity_SourceSeriesID_LongDesc]
      ,ImEx
      ,[DS Type] AS [ERSCommodity_DataSeriesCategory_Desc]
```

```
      ,12 --ERS
FROM [StagingISD].[dbo].[APDataSeries] A
LEFT JOIN MasterCoSD.CoSD.ERSCommoditySubCommodity_LU B ON
A.Commodity=B.ERSCommoditySubCommodity_Desc
```

To select Commodities that are not in Data Series

```
SELECT DISTINCT A.ERSCommoditySubCommodity_ID,A.ERSCommoditySubCommodity_Desc
  FROM MasterCoSD.CoSD.ERSCommoditySubCommodity_LU A
  WHERE A.ERSCommoditySubCommodity_GroupID IN(2,3,4)
  AND A.ERSCommoditySubCommodity_ID NOT IN (
  SELECT DISTINCT B.ERSCommoditySubCommodity_ID
  FROM AnimalProductsCoSD.CoSD.ERSCommodityDataSeries B)
```

To get count of each commodity in DataSeries

```
SELECT ERSCommoditySubCommodity_ID, count(ERSCommoditySubCommodity_ID)
  FROM AnimalProductsCoSD.CoSD.ERSCommodityDataSeries
  GROUP BY ERSCommoditySubCommodity_ID
  ORDER BY count(ERSCommoditySubCommodity_ID)
```

To get DataSeries not in Data Values:

```
SELECT DISTINCT A.ERSCommodity_ID
  FROM AnimalProductsCoSD.[CoSD].ERSCommodityDataSeries A
  WHERE A.ERSCommodity_ID NOT IN (
  SELECT DISTINCT B.ERSDataValues_ERSCommodity_ID
  FROM AnimalProductsCoSD.[CoSD].ERSDataValues B
  WHERE A.ERSCommodity_ID=B.ERSDataValues_ERSCommodity_ID)
```

To get DataValues not in Series

```
SELECT DISTINCT A.ERSDataValues_ERSCommodity_ID
  FROM AnimalProductsCoSD.[CoSD].ERSDataValues A
  WHERE A.ERSDataValues_ERSCommodity_ID NOT IN (
  SELECT DISTINCT B.ERSCommodity_ID FROM
  AnimalProductsCoSD.[CoSD].ERSCommodityDataSeries B
  WHERE A.ERSDataValues_ERSCommodity_ID=B.ERSCommodity_ID)
```

To split the Geography long description:

```
--------------Country----------------------
UPDATE [MasterCoSD].[CoSD].[ERSGeographyDimension_LU]
SET ERSGeographyDimension_Country = (substring( LEFT(ERSGeographyDimension_Desc,charindex(',
S: ',ERSGeographyDimension_Desc)-1),
       charindex('C: ',ERSGeographyDimension_Desc)+3,
```

```
	len(LEFT(ERSGeographyDimension_Desc,charindex(', S: ',ERSGeographyDimension_Desc)-1))-1
))


------------State------------
UPDATE [MasterCoSD].[CoSD].[ERSGeographyDimension_LU]
SET ERSGeographyDimension_State = (substring( LEFT(ERSGeographyDimension_Desc,charindex(', Co:
',ERSGeographyDimension_Desc)-1),
	charindex('S: ',ERSGeographyDimension_Desc)+3,
	len(LEFT(ERSGeographyDimension_Desc,charindex(', Co: ',ERSGeographyDimension_Desc)-1))-
1 ))


------------County------------
UPDATE [MasterCoSD].[CoSD].[ERSGeographyDimension_LU]
SET ERSGeographyDimension_County = (substring( LEFT(ERSGeographyDimension_Desc,charindex(', R:
',ERSGeographyDimension_Desc)-1),
	charindex('Co: ',ERSGeographyDimension_Desc)+3,
	len(LEFT(ERSGeographyDimension_Desc,charindex(', R: ',ERSGeographyDimension_Desc)-1))-1
))


------------Region------------
UPDATE [MasterCoSD].[CoSD].[ERSGeographyDimension_LU]
SET ERSGeographyDimension_Region = (substring( LEFT(ERSGeographyDimension_Desc,charindex(',
Ci: ',ERSGeographyDimension_Desc)-1),
	charindex('R: ',ERSGeographyDimension_Desc)+3,
	len(LEFT(ERSGeographyDimension_Desc,charindex(', Ci: ',ERSGeographyDimension_Desc)-1))-
1 ))


------------City------------
UPDATE [MasterCoSD].[CoSD].[ERSGeographyDimension_LU]
SET ERSGeographyDimension_City = (SUBSTRING(ERSGeographyDimension_Desc, CHARINDEX('Ci: ',
ERSGeographyDimension_Desc)+3, LEN(ERSGeographyDimension_Desc)))
```

Data Value migration for Veg Trade:

```
SELECT
	CASE
	    WHEN A.[Attrib] LIKE '%imports%' THEN '4'
	    WHEN A.[Attrib] LIKE '%exports%' THEN '3'
	END,
	9,
	F.ERSStatisticType_ID,
	NULL,
	NULL,
	H.ERSTimeDimension_ID,
```

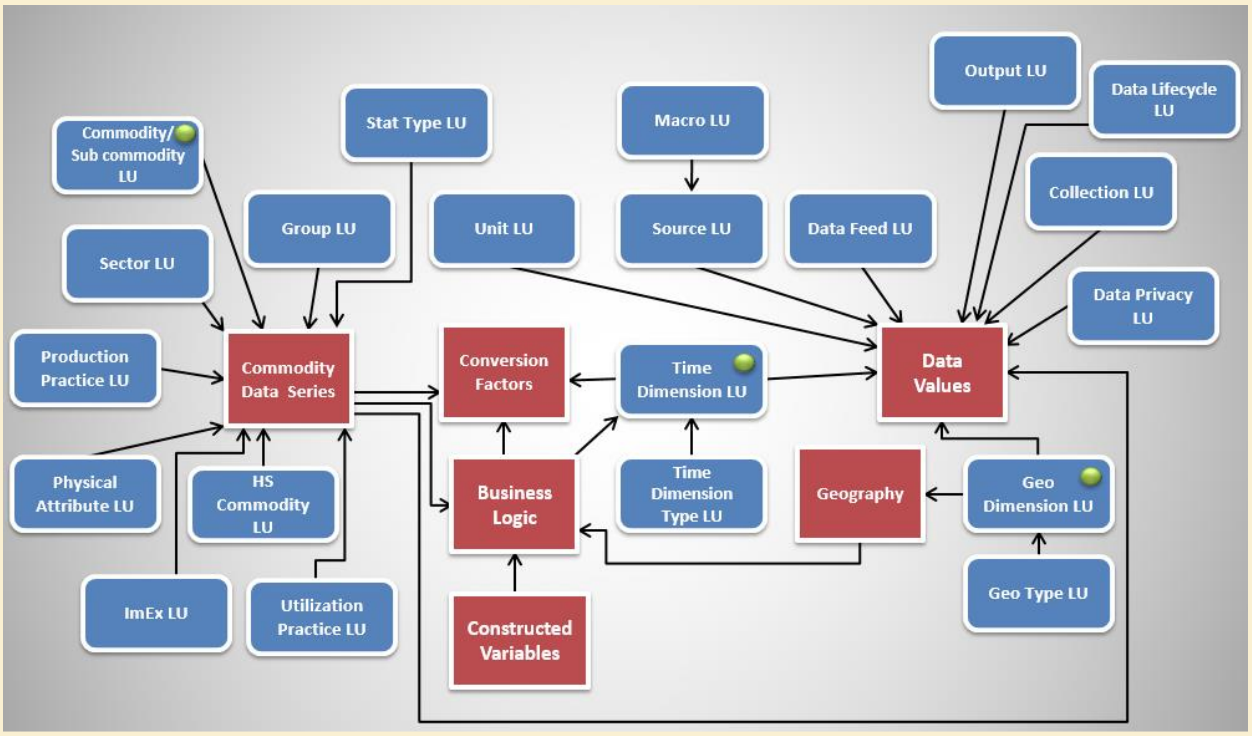
```
        G.ERSCommodity_ID,
        5,
        B.ERSUnit_ID,
        NULL,
        NULL,
        C.ERSGeographyDimension_ID,
        A.[Amount],
        NULL,
        NULL,
        4,
        6
FROM StagingISD.dbo.HS_Veg A
LEFT JOIN VegetablesCoSD.CoSD.ERSUnit_LU B ON A.UOM= B.ERSUnit_Desc
LEFT JOIN VegetablesCoSD.CoSD.ERSGeographyDimension_LU C ON C.ERSGeographyDimension_Country=
RIGHT(A.[Country], LEN(A.[Country]) - 7)
LEFT JOIN VegetablesCoSD.CoSD.ERSHS_LU D ON A.HSCode= D.ERSHS_Code
LEFT JOIN VegetablesCoSD.CoSD.ERSCommoditySubCommodity_LU E ON E.ERSCommoditySubCommodity_Desc
= D.ERSHS_Desc
LEFT JOIN VegetablesCoSD.CoSD.ERSStatisticType_LU F ON F.ERSStatisticType_Attribute=
RIGHT(A.Attrib, LEN(A.Attrib) - 15)
INNER JOIN VegetablesCoSD.CoSD.ERSCommodityDataSeries G ON A.HSCode=
G.ERSCommodity_SourceSeriesID
LEFT JOIN VegetablesCoSD.CoSD.ERSTimeDimension_LU H ON A.YearNum= H.ERSTimeDimension_Year
              AND ISNUMERIC(H.ERSTimeDimension_Month) = 1
              AND H.ERSTimeDimension_Month = A.MonthNum
              AND H.ERSTimeDimension_TimeDimensionType_ID=11
```

FILTERING AND STRUCTURE CHANGES

DATA STREAMING (FROM DIFFERENT SOURCES)

OVERALL DATABASE STRUCTURE

The CoSDs were implemented with 18 Lookup tables, 3 Aggregation tables, and 6 Data tables.



Lookup Tables:

ERSTimeDimensionType_LU:

	Column Name	Data Type	Allow Nulls
🔑	ERSTimeDimensionType_ID	int	☐
	ERSTimeDimensionType_Desc	varchar(100)	☐
	ERSTimeDimensionType_LongDesc	varchar(500)	☑

ERSGeographyType_LU

	Column Name	Data Type	Allow Nulls
🔑	ERSGeographyType_ID	int	☐
	ERSGeographyType_Desc	varchar(100)	☐
	ERSGeographyType_LongDesc	varchar(500)	☑

ERSCommoditySubCommodity_LU

Column Name	Data Type	Allow Nulls
ERSCommoditySubCommodity_ID	int	☐
ERSCommoditySubCommodity_Desc	varchar(100)	☐
ERSCommoditySubCommodity_IsMainMapping	varchar(50)	☐
ERSCommoditySubCommodity_LongDesc	varchar(500)	☑
ERSCommoditySubCommodity_GroupID	int	☐
ERSCommoditySubCommodity_Hierarchy	hierarchyid	☑
		☐

ERSTimeDimension_LU

Column Name	Data Type	Allow Nulls
ERSTimeDimension_ID	int	☐
ERSTimeDimension_TimeDimensionType_ID	int	☐
ERSTimeDimension_Desc	varchar(500)	☐
ERSTimeDimension_Date	date	☐
ERSTimeDimension_Year	int	☑
ERSTimeDimension_Month	int	☑
ERSTimeDimension_Day	int	☑

ERSGeographyDimension_LU

Column Name	Data Type	Allow Nulls
ERSGeographyDimension_ID	int	☐
ERSGeographyDimension_ERSGeographyType_ID	int	☐
ERSGeographyDimension_Desc	varchar(500)	☐
ERSGeographyDimension_Country	varchar(50)	☑
ERSGeographyDimension_State	varchar(50)	☑
ERSGeographyDimension_County	varchar(50)	☑
ERSGeographyDimension_Region	varchar(50)	☑
ERSGeographyDimension_City	varchar(50)	☑

ERSSector_LU

Column Name	Data Type	Allow Nulls
ERSSector_ID	int	☐
ERSSector_Desc	varchar(100)	☐
ERSSector_LongDesc	varchar(500)	☑

ERSCollection_LU

Column Name	Data Type	Allow Nulls
ERSCollection_ID	int	☐
ERSCollection_Desc	varchar(100)	☐
ERSCollection_LongDesc	varchar(500)	☑

ERSGroup_LU

Column Name	Data Type	Allow Nulls
ERSGroup_ID	int	☐
ERSGroup_Desc	varchar(100)	☐
ERSGroup_LongDesc	varchar(500)	☑
ERSGroup_POC_Name	varchar(100)	☐
ERSGroup_POC_EMAIL	varchar(100)	☑

ERSDataFeedType_LU

Column Name	Data Type	Allow Nulls
ERSDataFeedType_ID	int	☐
ERSDataFeedType_Desc	varchar(100)	☐
ERSDataFeedType_LongDesc	varchar(500)	☑
ERSDataFeedType_UpdateFrequency	varchar(100)	☑

ERSDataLifecycle_LU

Column Name	Data Type	Allow Nulls
ERSDataLifecyclePhase_ID	int	☐
ERSDataLifecyclePhase_Desc	varchar(100)	☐
ERSDataLifecyclePhase_LongDesc	varchar(500)	☑

ERSDataPrivacy_LU

Column Name	Data Type	Allow Nulls
ERSDataPrivacy_ID	int	☐
ERSDataPrivacy_Desc	varchar(50)	☐
ERSDataPrivacy_LongDesc	varchar(500)	☑

ERSGeographyCodes_LU

Column Name	Data Type	Allow Nulls
ERSGeography_ID	int	☐
ERSGeography_Desc	varchar(100)	☐
ERSGeography_LongDesc	varchar(500)	☑
ERSGeography_PSD_Code	varchar(10)	☑
ERSGeography_UST_Code	varchar(10)	☑
ERSGeography_State_ANSI_Code	varchar(10)	☑
ERSGeography_State_FIPS_Code	varchar(10)	☑
ERSGeography_County_Code	varchar(10)	☑
ERSGeography_Zip_Code	varchar(5)	☑
ERSGeography_ZipPlus4_Code	varchar(9)	☑
ERSGeography_ERSGeographyDimension_ID	int	☐
ERSGeography_LatitudeDD	float	☑
ERSGeography_LongitudeDD	float	☑
ERSGeography_LatitudeDMS	float	☑
ERSGeography_LongitudeDMS	float	☑
ERSGeography_Country_Code	varchar(10)	☑
ERSGeography_County_ANSI_Code	varchar(10)	☑

ERSHS_LU

Column Name	Data Type	Allow Nulls
ERSHS_ID	int	☐
ERSHS_Desc	varchar(100)	☐
ERSHS_LongDesc	varchar(500)	☑
ERSHS_Code	varchar(50)	☐

ERSImEx_LU

Column Name	Data Type	Allow Nulls
ERSTradeImEx_ID	int	☐
ERSTradeHS10_ImportExport	varchar(50)	☐

ERSOutput_LU

Column Name	Data Type	Allow Nulls
ERSOutput_ID	int	☐
ERSOutput_Desc	varchar(100)	☐
ERSOutput_LongDesc	varchar(500)	☑
ERSOutput_POC_Name	varchar(100)	☑

ERSPhysicalAttribute_LU

Column Name	Data Type	Allow Nulls
ERSPhysicalAttribute_ID	int	☐
ERSPhysicalAttribute_Desc	varchar(100)	☐
ERSPhysicalAttribute_LongDesc	varchar(500)	☑

ERSProdPractice_LU

Column Name	Data Type	Allow Nulls
ERSProdPractice_ID	int	☐
ERSProdPractice_Desc	varchar(100)	☐
ERSProdPractice_LongDesc	varchar(500)	☑

ERSStatisticType_LU

Column Name	Data Type	Allow Nulls
ERSStatisticType_ID	int	☐
ERSStatisticType_Attribute	varchar(100)	☐
ERSStatisticType_Attribute_Desc	varchar(100)	☐
ERSStatisticType_Attribute_LongDesc	varchar(500)	☑
ERSStatisticType_Mapping	varchar(50)	☐

ERSTool_ActionLog

Column Name	Data Type	Allow Nulls
ERSToolActionLog_ID	int	☐
ERSToolActionLog_time	datetime	☑
ERSToolActionLog_User	varchar(500)	☑
ERSToolActionLog_Desc	varchar(MAX)	☑

ERSUnit_LU

Column Name	Data Type	Allow Nulls
ERSUnit_ID	int	☐
ERSUnit_Desc	varchar(100)	☐
ERSUnit_LongDesc	varchar(500)	☑
ERSUnit_Type	varchar(50)	☑

ERSUtilPractice_LU

Column Name	Data Type	Allow Nulls
ERSUtilPractice_ID	int	☐
ERSUtilPractice_Desc	varchar(100)	☐
ERSUtilPractice_LongDesc	varchar(500)	☑
ERSUtilPractice_NAIC_Code	int	☑

DATA VALIDATION AND VERIFICATION

BUSINESS RULES

THE COSD TOOL

VISUALIZATIONS AND PIVOT TABLES WITH COSD

DATA LIFECYCLE

COSD DATA PRIVACY

MASTER TAXONOMY

DATA REPOSITORY