Combined Air Emissions Short-Term Wins
Data Dictionary Project Description

1. Project Name: Harmonization of Cross-Program Code Tables, Data Elements

2. Project Participants: *Note roles where applicable.*
  John Harman, OEI-OIC, Co-Lead
  Tammy Manning, North Carolina, Co-Lead
  Lauren Gordon, OEI-OIC, Member
  Matt Kelly, OEI-OIC, Member
  Theresa Lowe, OAR-OAQPS-SPPD-MPG, Member
  Josh Drukenbrod, OAR-OAQPS-AQAD-EIAG, Member
  Mark Wert, Massachusetts, Member

3. Project Funding: $70K

Contractor support for:
- Processing and updating data dictionaries
  - Ensuring clear names and definitions (to allow for cross-walking)
- Comparing Data Elements
  - Identifying data elements that have the same meaning and could be standardized across data dictionaries
  - Identifying data elements that sound the same but have different meanings
  - Identifying data elements that are similar and potentially could be standardized
  - Creating a map across programs of data elements shared by 2 or more programs
- Processing and updating code sets contained in data dictionaries
  - Ensuring clear names and definitions (to allow for cross-walking)
- Comparing code set values
  - Identifying code set values that have the same meaning and could be standardized across data dictionaries
  - Identifying code set values that sound the same but have different meanings
  - Identifying code set values that are similar and potentially could be standardized
  - Creating a map across programs of data elements shared by 2 or more programs
- Prepare material for explaining results
- Provide current and harmonized data elements and code sets to system owners for implementation in systems

4. How project fits within the larger Combined Air Emissions "to be" state
The final mapping report will show how the system data elements map to one another to allow for designing a portal that provides data to all programs (and reduced redundant data collection).

In the "to be" state, inventory data comes in once and is distributed to multiple air agencies. To do this, the code tables need to be harmonized across the multiple programs where the same types of codes are needed (for example: facility status code of "operating' may mean

"emitting" for one program and "permitted" for another program). In addition, for data to come in through a portal once and be distributed, the fields coming in from the portal must be mapped to the multiple data systems that will want some data but not all.

5. Goals:
- Goal 1: Create actual code tables that could be used in the "to be" state. Expected to be in spreadsheets with accompanying report explaining what was done.
- Goal 2: Create a map of data elements for those elements that are in common across the EPA, state, local, and tribal air data systems. This will ideally include all systems, but resources may not allow that; creating and mapping one or two state data dictionaries would provide an important base from which to build further assessments. Expected to be a spreadsheet or database.

6. Project description:
Tasks
1. Work across all inventory development groups to compare and harmonize code tables.
2. Reduce and eliminate differences of code sets where possible to reduce LOE for SLTs and facilities, and support concept of future single submissions.
3. Create a product that shows the mapping of data elements, and code set values, across systems/programs. This will support the new FRS data model and future single-submission concepts. Note: previously called 'Increase/improve communication among internal EPA databases and program staff.'

7. Key Challenges:
- **Challenge 1**. Obtaining NEI data dictionaries for two states and processing them (ensuring clear names and definitions and cross-walking with other data dictionaries).
  - o Work with ECOS to obtain state data dictionaries
- **Challenge 2**. Determining similar but not identical data elements and code set values.
  - o System/program owners must agree on the definitions proposed by the workgroup and reach consensus on degree to which these can be harmonized
- **Challenge 3**. For system/program owners, there may be a need to update systems.
  - o When the workgroup reaches consensus on the data elements and code set values that are the same, system/program owners must plan to update their systems to reflect the decisions.
- **Challenge 4**. For system/program owners, there may be a need to update regulations and Information Collection Requests (ICRs).
  - o When the workgroup reaches agreement that two or more data elements are similar and could be modified to be identical, these changes in definition may indicate whether there is a need to revise regulations or ICRs.

8. Expected workload for team participants:
- Short bi-weekly meetings to provide status of contractor effort.
- Bi-weekly meetings for workgroup to agree on findings from contractor effort and reach consensus on those data elements and code set values that have the same definitions or that are similar and could be harmonized.

9. Quality Assurance: *If applicable, list quality assurance issues that might arise during this project and how quality review will be included to ensure the best outcomes. Add work products if needed to ensure any potential quality problems will be addressed.*

10. Project tasks, work products, costs, and expected completion dates:

Task	Work product	Expected Task Cost	Expected end date	Actual End Date
Task 1: Data Dictionary Work*		51,000$	MM/DD/YY	MM/DD/YY**
	Processed CEDRI data dictionary		10/31/15	12/31/15
	Processed EIS data dictionary		10/31/15	10/31/15
	Processed TRI data dictionary		10/31/15	12/31/15
	Processed GHG data dictionary		10/31/15	1/31/16
	Processed IOWA NEI data dictionary		10/31/15	1/31/16
	Processed North Carolina AERO NEI data dictionary		10/31/15	1/31/16
	Processed North Carolina ED NEI data dictionary		10/31/15	1/31/16
	Processed Texas NEI data dictionary		10/31/15	1/31/16
Task 2: Code Set Work		12,000$		
	Processed CEDRI code set values		10/31/15	12/31/15
	Processed EIS code set values		10/31/15	10/31/15
	Processed TRI code set values		10/31/15	1/31/16
	Processed GHG code set values		10/31/15	1/31/16
	Processed Iowa NEI code set values		10/31/15	TBD –not received
	Processed North Carolina AERO NEI code set values		10/31/15	1/31/16
	Processed North Carolina ED NEI code set values		10/31/15	1/31/16
	Processed Texas NEI code set values		10/31/15	1/31/16
Task 3: Comparison Work		7,000		
	Draft findings comparing data elements		12/31/15	12/31/15
	Draft findings comparing code set values		12/31/15	12/31/15
	Draft report comparing data elements		2/29/16	
	Draft report comparing code set values		2/29/16	
	Presentations to subgroup members, ECOS, and invited guests		3/31/16	
	Final reports		3/31/16	
	Presentations to full workgroup using info in final reports		4/30/16	
Task 4: Workgroup Harmonization				
	Harmonize identical data elements		TBD	
	Harmonize identical code set values		TBD	

Task	Work product	Expected Task Cost	Expected end date	Actual End Date
	Establish similar data elements and code set values for possible programmatic or regulatory changes		TBD	
TOTAL COST:		70,000$		

*Data dictionary work builds on data dictionary processing already performed.

**Final data dictionary work may have late complete dates because of editorial QC reviews.