1. **The document provided to you regarding the roles and responsibilities of the CDO**



1. **Data Sharing agreements, data elements we have already gathered and the pursuit of our “data inventory”.  (for the latter, the categories of data and our approach)**

**Commonwealth Master Data Agreement (MDA)**

Commonwealth of Kentucky executive cabinets have been collecting “business data” in various formats for a longtime as part of normal operations. This data drives day-to-day activities of the functional areas and their business operations in support of citizen services but also have enormous value in addressing Commonwealth’s most significant challenges and can support data-driven policy decision making.

Many government agencies have been sharing data in various forms for a long time. However, each occurrence generally requires a data use agreement of some sort, which will have to go through rigorous reviews and approval process by legal teams on both ends. This entire process may have to be repeated when another distinct set of data are to be exchanged between the same parties. It is also hard to keep track of those final agreements, changes to each agreement, maintaining a repository of those agreements, audits, mapping security dependencies etc.

To make it easy for Commonwealth agencies to share data for appropriate use cases in a timely manner and streamline the entire process, Commonwealth Chief Compliance Office has developed a Master Data Agreement (MDA). The following are the salient characteristics of this MDA:

* Only the MDA requires the Secretary’s signature.  It is labeled Master Commonwealth Data Agreement at the bottom of the page.
* **Schedule A** requires completion of the Point of Contact for each Cabinet, the copy attached has all the contacts proved to date.  The Cabinet does not have to provide this role at signing if it is not decided upon.  Labeled Master Commonwealth Data Agreement, Schedule A
* **Schedule B** requires no signature, is informational concerning the duties of the Cabinet Point of Contact. Labeled Master Commonwealth Data Agreement, Schedule B
* **Schedule C** requires no signature, is informational concerning the duties of the Cabinet Point of Contact. Labeled Master Commonwealth Data Agreement, Schedule C
* **Schedule D or Schedule** E is required when the Cabinet decides to share a set of data with another Cabinet. Each instance need to be documented using Schedule D (State owned Data) or Schedule E (Federally Regulated Data). These schedules do not require the Cabinet Secretary’s signature.  They can be signed by an executive with the authority to authorize the sharing of the data described.  This could be the individual designated as the Point of Contact (POC).  However, each Cabinet will have to determine who is required to sign these documents.  These schedules do not have to be completed until an agency decides to share data with another agency.
* **Schedules F1, F2 and G** are informational and require no signature

The following is a list of Cabinets that have signed the Commonwealth of Kentucky Master Data Agreement (MDA) thus far.

1. Cabinet for Health and Family Services
2. Justice and Public Safety Cabinet
3. Finance and Administration Cabinet
4. Labor Cabinet
5. State CIO



**Metadata Inventory:**

Our metadata inventory approach categorizes the Commonwealth of Kentucky business data into two board groups –

1. Person-based (people)
2. Entity-based (businesses)

Most business applications in the Commonwealth falls into one of these categories. Some of these business applications are the primary source for a specific set of data about people or entities. Identifying these “Golden Record” systems (aka, System of Record) would be extremely important to establish a baseline for authoritative data sources, which could enable deep data analytics in support of sound policy decisions. In addition, these authoritative systems can reduce the data duplication and redundancies while providing high quality data lookups/referencing across the agencies. The Citizen or Business interactions with the State government can be turned into a great user experience by promoting “Collect once and reuse” approach across the Commonwealth of Kentucky. However, it is vital to assure the quality of data close to 100% at every stage of collection in the “Golden Record” systems.

Per Dr. Grindle’s direction, we have been working on documenting enterprise-wide high value data elements/attributes. The goal of this effort is to derive better value from the data assets we already own for government operations and citizen services. This “metadata repository” will enable quick responses to business questions from the Legislature, Governor’s office, Cabinet leadership, and the CIO. By reducing the barriers that impede the correlation of information in disparate systems, we can utilize this metadata as a more effective tool to provide solutions to many social issues.

Agency Data source metadata inventory reveals –

* 1. Types of data business accumulate/use/distribute/report on
  2. Identify Intra-Cabinet (within the Cabinet) Data duplication

The initial metadata collection consists of three templates as listed below:

1. Data Sources
2. Datasets/Interfaces
3. High-Value (HV) Data Elements
4. **Data Sources**

Generally, we can equate a **Business Application** with a **Data Source** but sometimes one Business Application may have multiple “modules” or “databases/schemas” so to speak. Data Sources are expected to be at the lowest granularity possible – say, a database.

1. **Datasets (aka Interfaces)**

List of all inbound and outbound datasets/interfaces (along with attributes shared in each interface) in and out of agency Business Applications.

1. **High-Value (HV) Data Elements**

This inventory will help with Data sharing across the agencies and Data Analytics in support of policy decisions within the cabinet and across government agencies. We have provided about 41 Business Applications based on our research at the time but we have requested all executive agencies to identify HV data elements from within their cabinet and add to it so we can grow the HV repository. We are also hoping to use this HV repository to identify “Golden” record systems within the state government. The initially identified data sources are mostly “person” based but this repository shall include “entities” too.

Though this is an ongoing effort, we have documented about 60 High-Value (HV) business data sources and several attributes from each of those business applications so far. However, these functional data elements will have to be “mapped” to the real physical databases and columns so that the Data Analytics teams can request for this data by direct reference to appropriate business applications and database names when needed. Agency subject matter experts (SME) and application support teams with business knowledge can only accurately do the effective mapping of these functional data elements to the actual databases.



1. **On Decision Lens, any legislation which has been passed in other states supporting open data sharing and the environments we have demo’ed (Socrata and CKAN Open Data Portal) that we are going to review.**

**Decision Lens**

Decision Lens is an Enterprise software for prioritizing highest value projects and eliminate wasted limited resources. COT has already invested in this tool and using it to prioritize Capital Projects for next fiscal year.

**Data Portal**

To enable Data sharing across Commonwealth agencies internally and to promote public government transparency, we have looked at couple of options for data portals.

1. Open source software – CKAN data portal
2. Socrata Connected Government Cloud (SCGC)

While the Internal portal will enable standardized and secure data sharing platform for state agencies, the Public-facing Open Data Portal will serve publication of machine-readable datasets and provide a single standardized platform for citizens and researchers. We are not alone on this as most other States already have statewide internal and open data portals.

**CKAN POC:**

The chief advantages of CKAN includes quicker new features addition when available, low TCO, and can be hosted onsite. However, we will have to develop internal skills to care-and-feed for this portal. A POC of CKAN has been completed successfully.

**Socrata Demo:**

Major of the States and large counties across the country have been using this platform due to its ease of use and standardized dataset delivery. COT has scheduled a demo of this product for all Data Analytics Working Group (DAWG) members and COT Executive staff later this month.

Date of Socrata demo:  Tuesday, March 26, 2019 10:00 AM-11:30 AM

**Open Data Portal Legislations – Quick Look**

There are a number of open data portal initiatives across the country. The National Conference of State Legislatures provides a summary ([here](http://www.ncsl.org/research/telecommunications-and-information-technology/open-data-legislation.aspx)). Many of the cited plans only touch on open data portals obliquely.

A number of cities have developed specific open data plans as well. NYC is highlighted below. Seattle is another exemplar and interestingly are pursuing Socrata (which we are reviewing) and a Data Maturity Model. Seattle has an [Open Data Plan](https://www.seattle.gov/Documents/Departments/Tech/OpenData/City_of_Seattle_2018_Open_Data_Plan.pdf) that provides a decent business case to support the policy and legislation. Louisville developed an open data plan in 2013 by [Executive Order](https://data.louisvilleky.gov/sites/default/files/Open-Data-Executive-Order-2013-signed.pdf). (The Louisville Open Data Portal is a good example of the impact that data can make if it is curated and published properly. Interestingly, it wasn’t clear that many of the pieces of legislation addressed the cost of data publication and curation.

Generally, these address all or some of the following issues:

* Open Data definitions and required security protections associated with managing, maintaining, and publishing. Should relate to our open data legislation (which might be worth revisiting in the light of the publication of open data via a portal).
* Organizational changes have generally been made in concert with the open data initiative. Specifically, the Chief Data Officer role and the Chief Privacy Officer role are created/addressed in numerous instances. Since we already have those roles, they would simply need to be reviewed in terms of their responsibilities for the inventory, management, and publication of open data. The newer legislative acts require data agents (what we are terming agency data stewards) for all agencies.
* Requirements to publish are generally required in the newest legislation. New York might be interesting since they introduce a unique three-tier system for prioritizing publication of data. All new applications / datasets are published by default (within 30 days of collection), legacy datasets have a negotiated timeframe and there is a class that fall in between (can’t be published in 30 days but are not legacy – which they mark as urgent and track their scheduling).
* Most require a data inventory/or catalog as part of the effort – similar to what we are trying to do administratively.
* Arkansas extended their open data effort to developing a business case for a statewide data warehouse and considering master data management. While a statewide data warehouse appears to be a simplistic technical solution, the practice (not technology) of master data management is an interesting adjunct to an open data portal.
* Finally, most of the new pieces of legislation have extended the portal to include state, regional, local, higher education and other organizations. A few have a review or advisory board although their role seems to be minor. Maryland is of interest in this area given how extensive they treated this board. They have also made procurement activities answerable to the open data legislation.

I have attached a bit of pertinent legislation. This has been a cursory review at best. A few specific legislative actions may be of interest. These were selected to illustrate the spectrum of legislation that has been sampled – not specific applicability to the Commonwealth. It would not be difficult to develop specific language for Kentucky. The long poles in the tent are likely 1) aligning the open data portal initiative to existing legislation and organizational responsibilities (particularly if Executive agencies are included), and 2) determining how to handle the increased cost for the service. Seattle estimated the expense for their Open Data Portal for one year was about $1M.

[Virginia](https://lis.virginia.gov/cgi-bin/legp604.exe?181+ful+CHAP0679+pdf) – In the Government Data Collections and Dissemination Practices Act requires a Chief Data Officer assigned to the Office of the Secretary of Administration to “oversee effective sharing of data among state, regional, and local public entities and public institutions of higher education and to implement effective data governance strategies to maintain data integrity and security and promote access to open data.

While there is a nod to open data, the legislation primarily focuses on setting up an administrative framework for managing data sharing.

[Florida](http://laws.flrules.org/2018/9) – In Chapter 2018-9, the State of Florida initiated their open portal initiative with a data inventory with particular attention to the publication of a machine readable open data catalog.

[State of New York](https://www.governor.ny.gov/news/no-95-using-technology-promote-transparency-improve-government-performance-and-enhance-citizen). The State of New York and the City of New York appear to have the most mature open data portal / program with associated legislation. All data are assumed to be public and open unless specifically excluded as supported by legislative definitions and restrictions. These laws address new data sets and legacy data sets and provide guidance regarding the urgency with which data must be made public.

NY State provides for an open data website.

1. An online open data website for the collection and public dissemination of publishable state data is hereby established in the office of information technology services. The open data website shall be maintained at data.ny.gov or such other successor website maintained by, or on behalf of, the state, as deemed appropriate by the office of information technology services. The open data website will provide access to publishable state data that is owned, controlled, collected or otherwise maintained by covered state entities.
2. The office of information technology services, in coordination with the committee for open government, and the data working group shall coordinate implementation and expansion of the open data website to facilitate the sharing of information.

[NY City](https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=3136796&GUID=0144E84F-0BEC-4DB9-ABC6-B564DE4F717C&Options=ID%7CText%7C&Search=%22open+data%22) appears to have a parallel structure, which is more robust. Of interest is a mandate for agencies to appoint data coordinators to administer and enforce the process.

[Arkansas House Bill 1793](http://www.arkleg.state.ar.us/assembly/2017/2017R/Bills/HB1793.pdf) provided for essentially the same basic instructions to provide and publish open data with the creation of a data inventory. Of some interest is the emphasis on data quality and master data management that this legislation introduces as well as language specific to business intelligence and data analysis. The instructions to the Department of Information Systems was to develop a business case for a statewide data warehouse with instructions for publishing data to the public.

\* \* \*