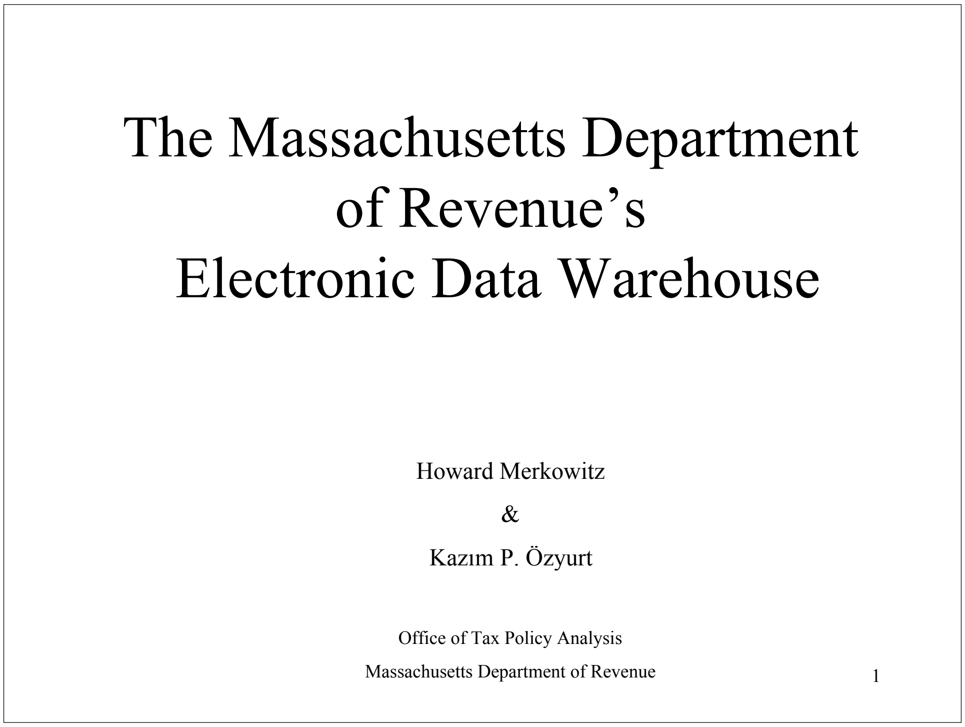


The Massachusetts Department of Revenue's Electronic Data Warehouse

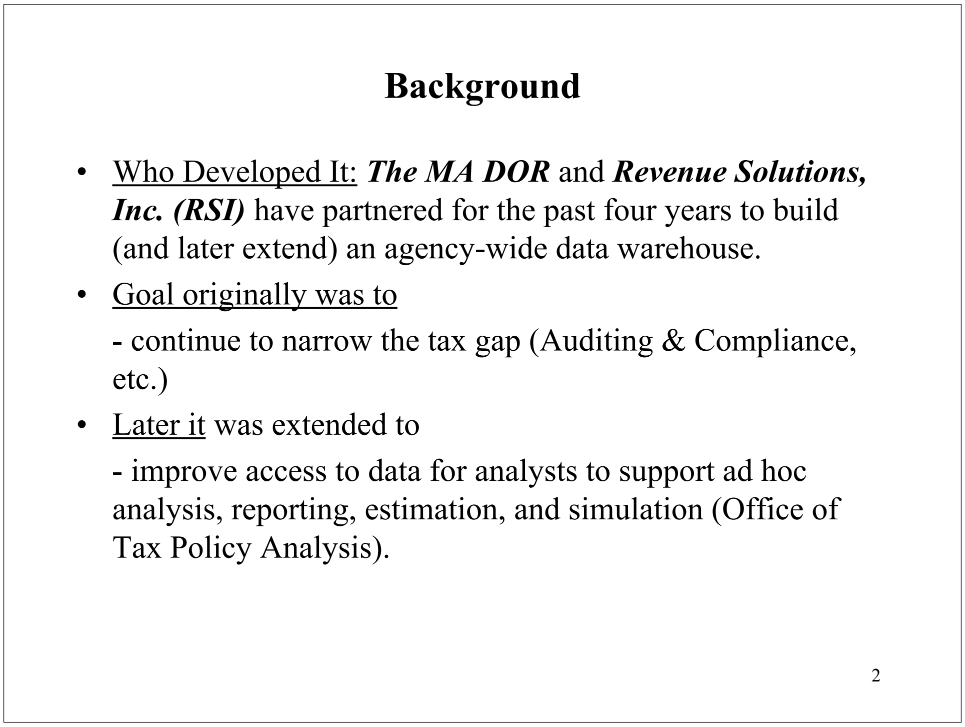
Howard Merkowitz

&

Kazım P. Özyurt

Office of Tax Policy Analysis

Massachusetts Department of Revenue

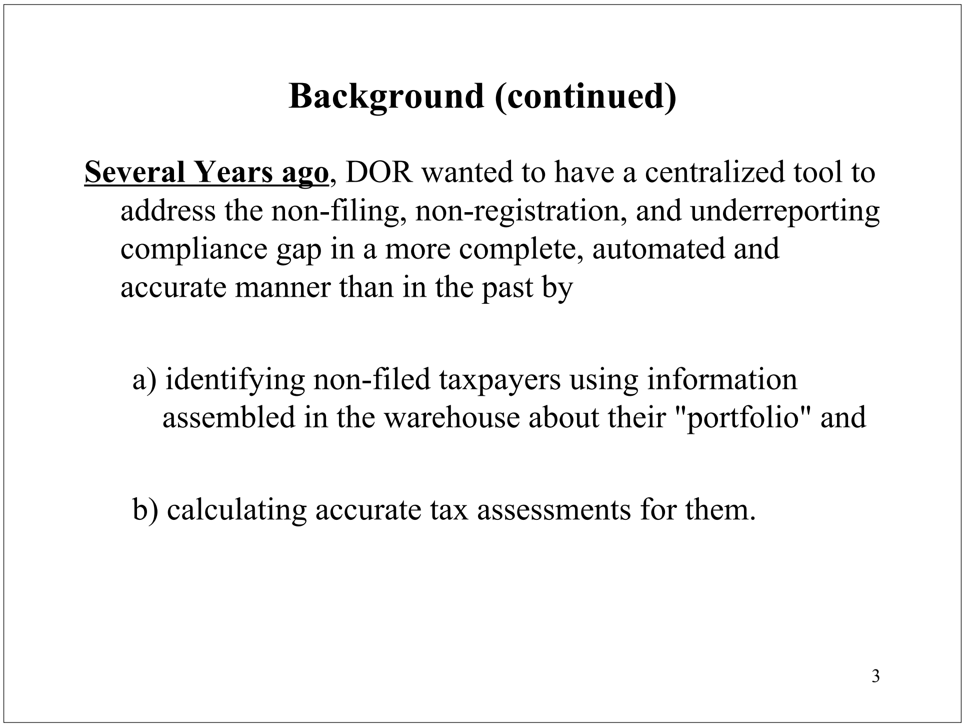


Background

- Who Developed It: ***The MA DOR*** and ***Revenue Solutions, Inc. (RSI)*** have partnered for the past four years to build (and later extend) an agency-wide data warehouse.
- Goal originally was to

  - continue to narrow the tax gap (Auditing & Compliance, etc.)
- Later it was extended to

  - improve access to data for analysts to support ad hoc analysis, reporting, estimation, and simulation (Office of Tax Policy Analysis).

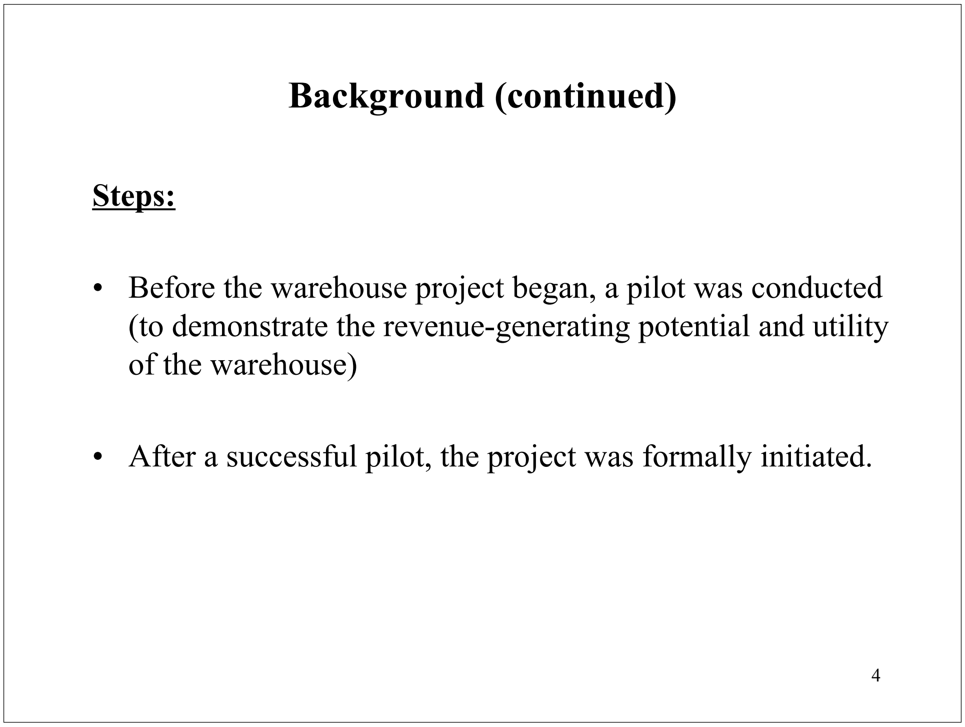


Background (continued)

Several Years ago, DOR wanted to have a centralized tool to address the non-filing, non-registration, and underreporting compliance gap in a more complete, automated and accurate manner than in the past by

a) identifying non-filed taxpayers using information assembled in the warehouse about their "portfolio" and

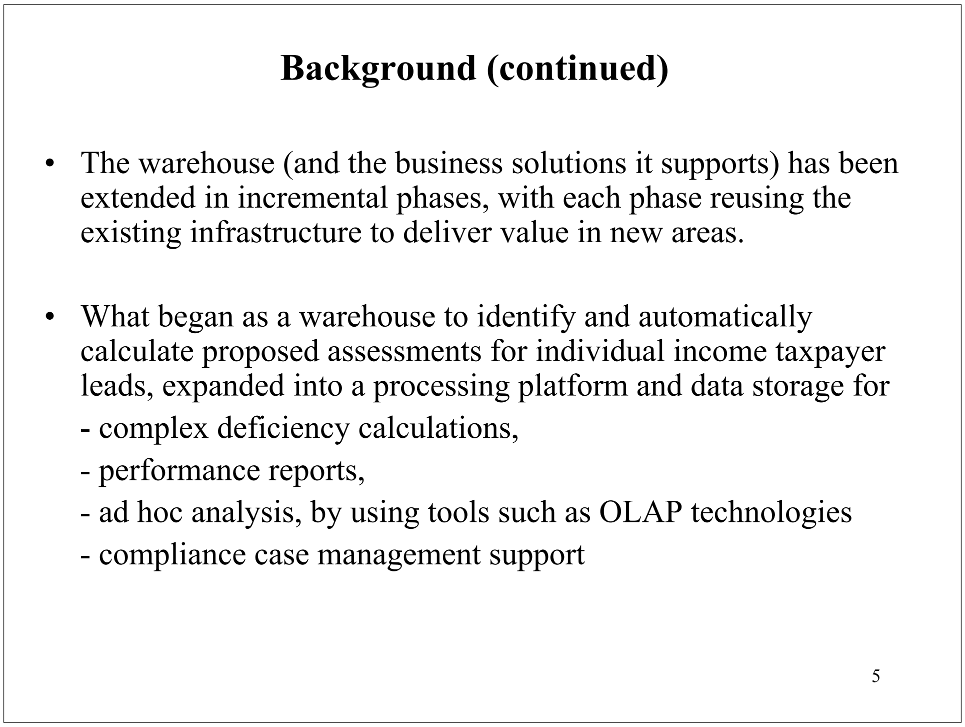
b) calculating accurate tax assessments for them.



Background (continued)

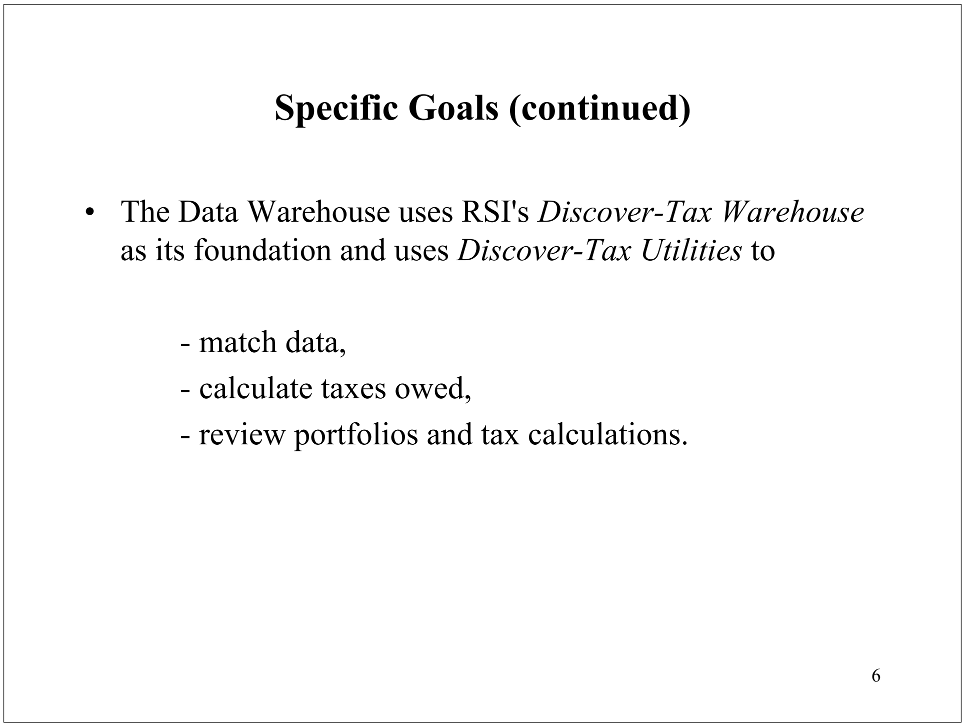
Steps:

- Before the warehouse project began, a pilot was conducted (to demonstrate the revenue-generating potential and utility of the warehouse)
- After a successful pilot, the project was formally initiated.



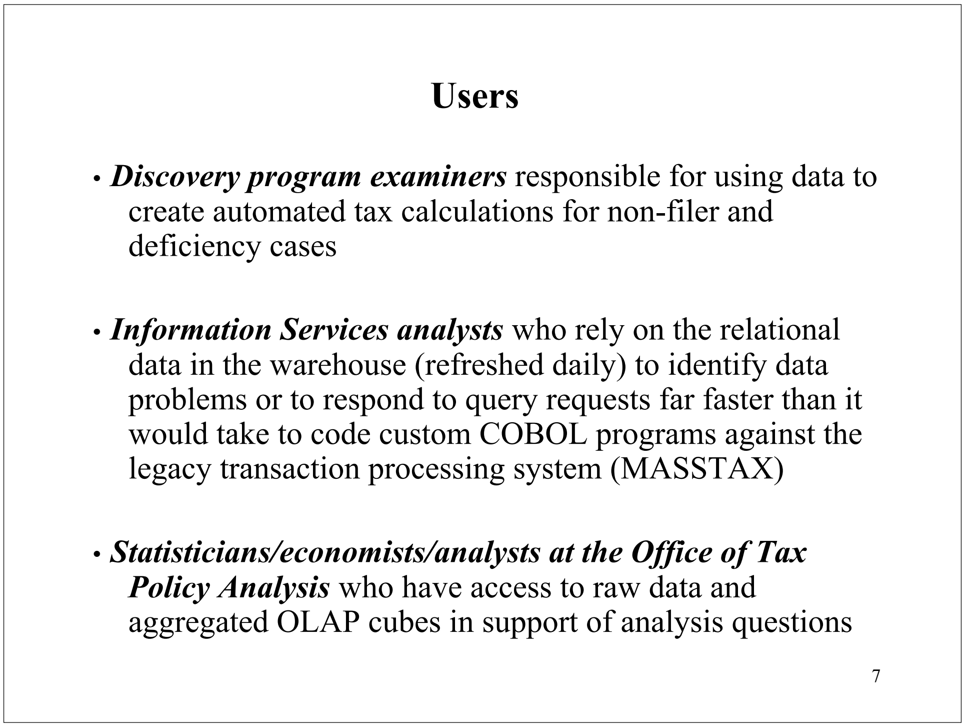
Background (continued)

- The warehouse (and the business solutions it supports) has been extended in incremental phases, with each phase reusing the existing infrastructure to deliver value in new areas.
- What began as a warehouse to identify and automatically calculate proposed assessments for individual income taxpayer leads, expanded into a processing platform and data storage for
  - complex deficiency calculations,
  - performance reports,
  - ad hoc analysis, by using tools such as OLAP technologies
  - compliance case management support



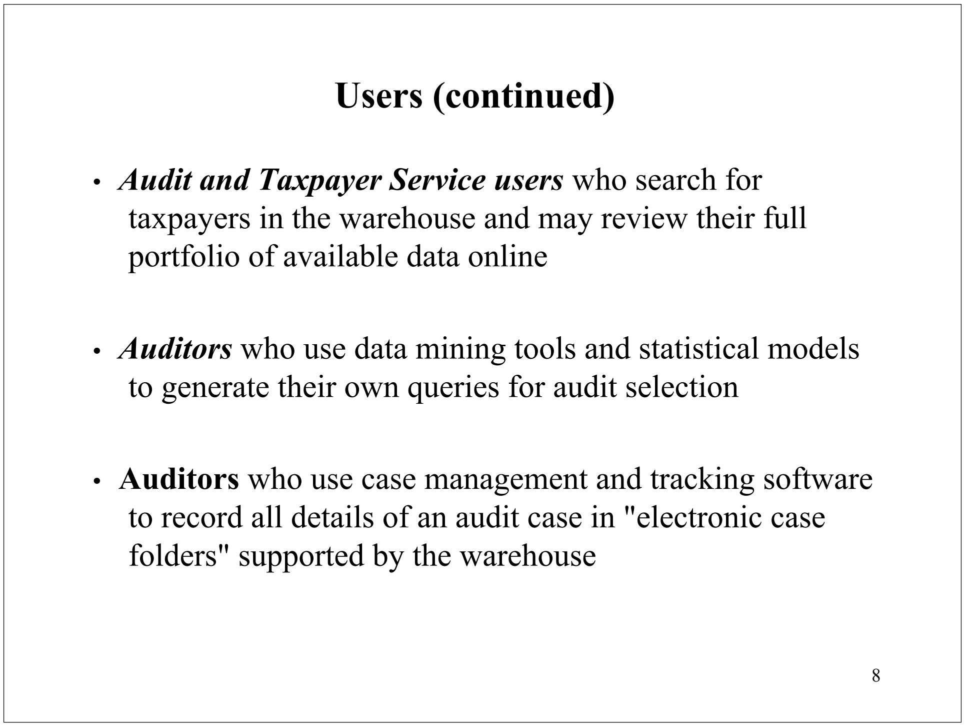
Specific Goals (continued)

- The Data Warehouse uses RSI's *Discover-Tax Warehouse* as its foundation and uses *Discover-Tax Utilities* to
  - match data,
  - calculate taxes owed,
  - review portfolios and tax calculations.



Users

- ***Discovery program examiners*** responsible for using data to create automated tax calculations for non-filer and deficiency cases
- ***Information Services analysts*** who rely on the relational data in the warehouse (refreshed daily) to identify data problems or to respond to query requests far faster than it would take to code custom COBOL programs against the legacy transaction processing system (MASSTAX)
- ***Statisticians/economists/analysts at the Office of Tax Policy Analysis*** who have access to raw data and aggregated OLAP cubes in support of analysis questions



Users (continued)

- ***Audit and Taxpayer Service users*** who search for taxpayers in the warehouse and may review their full portfolio of available data online
- ***Auditors*** who use data mining tools and statistical models to generate their own queries for audit selection
- **Auditors** who use case management and tracking software to record all details of an audit case in "electronic case folders" supported by the warehouse

Users (continued)

- **Collectors** who use specialized utilities in the warehouse to identify lien and levy sources, offset opportunities, and demographic data
- **Collections Management** who have implemented a sophisticated risk-based approach to assigning actions to different collections cases by using statistical models and strategy rules that are executed in the warehouse for new cases and which rely on historical taxpayer data in the warehouse.
- **Managers** who use an evolving executive-level reporting dashboard of key measures derived from the warehouse to monitor performance against their goals. (*Still under development.)*

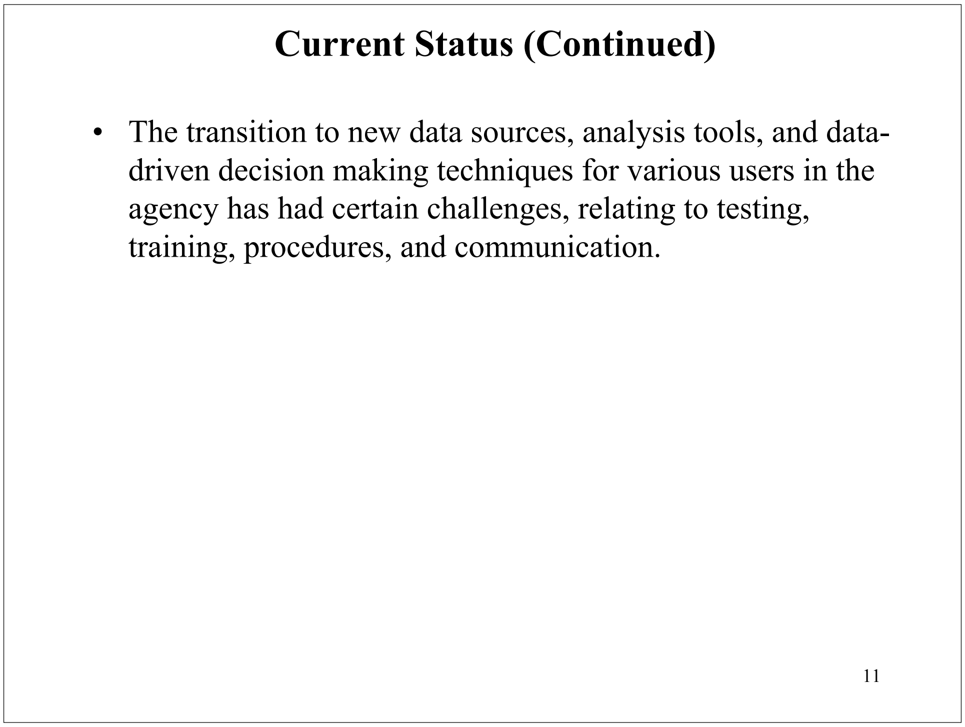
Current Status

- The warehouse has supported tax discovery programs that have generated significant *new revenues* for the Commonwealth - in excess of $325M.

- The warehouse has also begun to "unlock" data that is now *used by more groups* in the agency for reporting, analysis, and compliance management.

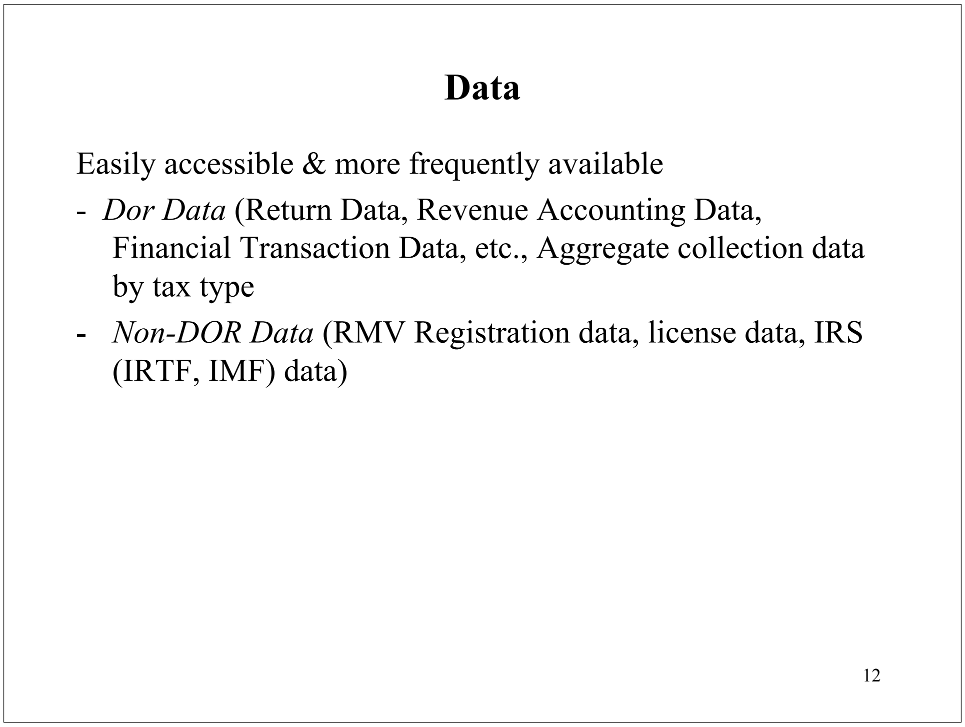
  - The *requests for additional* data have grown substantially.

  - The warehouse itself has *grown to over 5 terabytes* of data and includes several hundred data sources, with additional users coming with *additional needs* for more current and more varied sources of data.



Current Status (Continued)

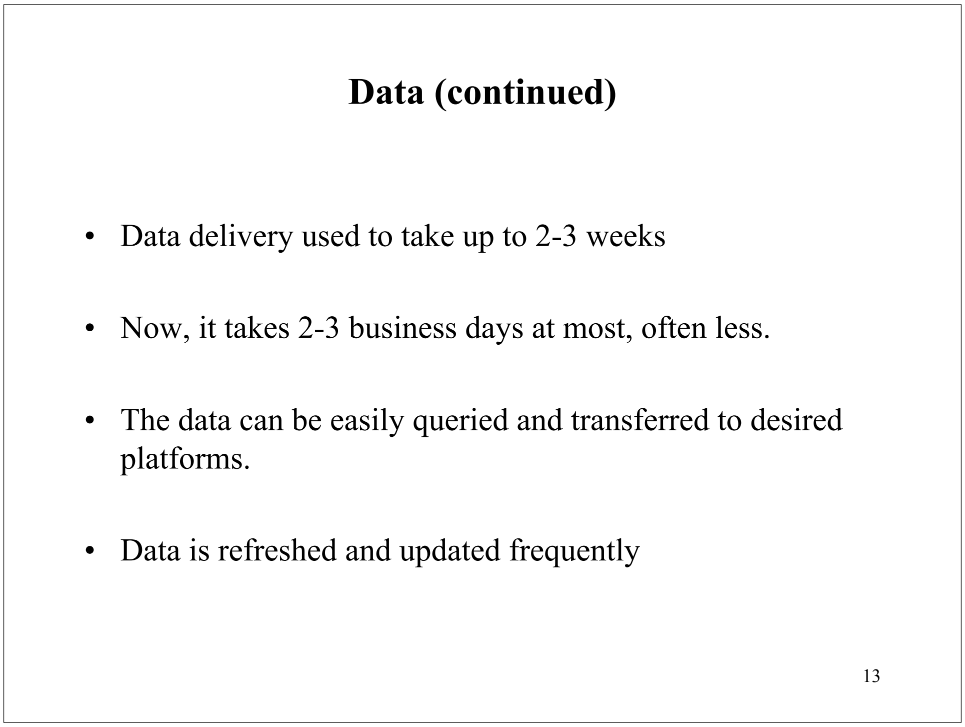
- The transition to new data sources, analysis tools, and data-driven decision making techniques for various users in the agency has had certain challenges, relating to testing, training, procedures, and communication.



Data

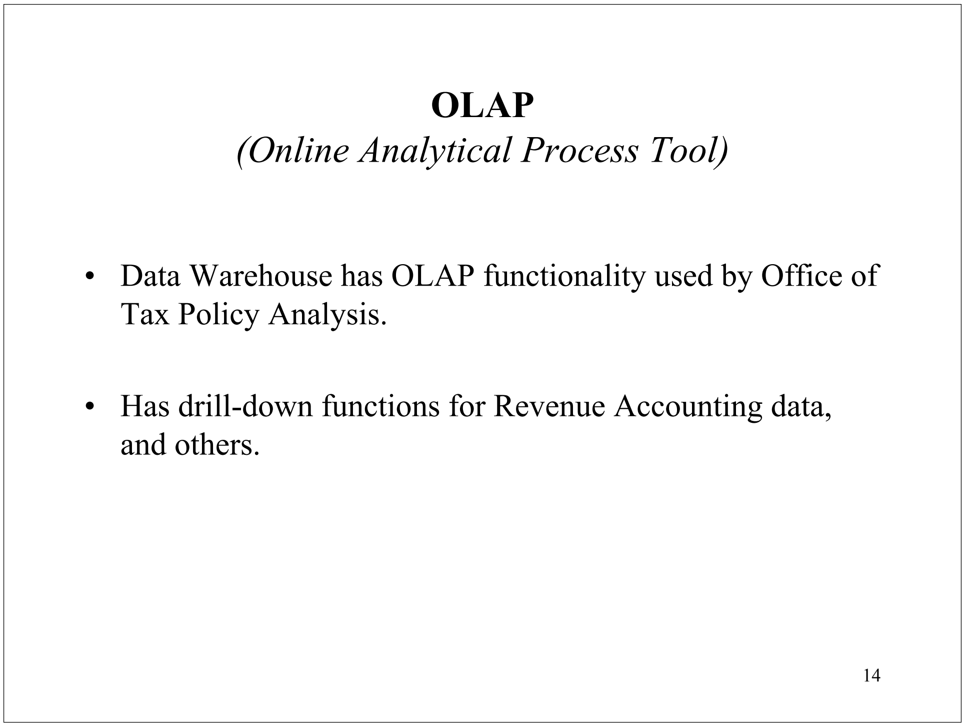
Easily accessible & more frequently available

- *Dor Data* (Return Data, Revenue Accounting Data, Financial Transaction Data, etc., Aggregate collection data by tax type
- *Non-DOR Data* (RMV Registration data, license data, IRS (IRTF, IMF) data)



Data (continued)

- Data delivery used to take up to 2-3 weeks
- Now, it takes 2-3 business days at most, often less.
- The data can be easily queried and transferred to desired platforms.
- Data is refreshed and updated frequently



OLAP
(Online Analytical Process Tool)

- Data Warehouse has OLAP functionality used by Office of Tax Policy Analysis.
- Has drill-down functions for Revenue Accounting data, and others.

Tax Calculator

(Under in development)

- Under development for running return data microsimulations

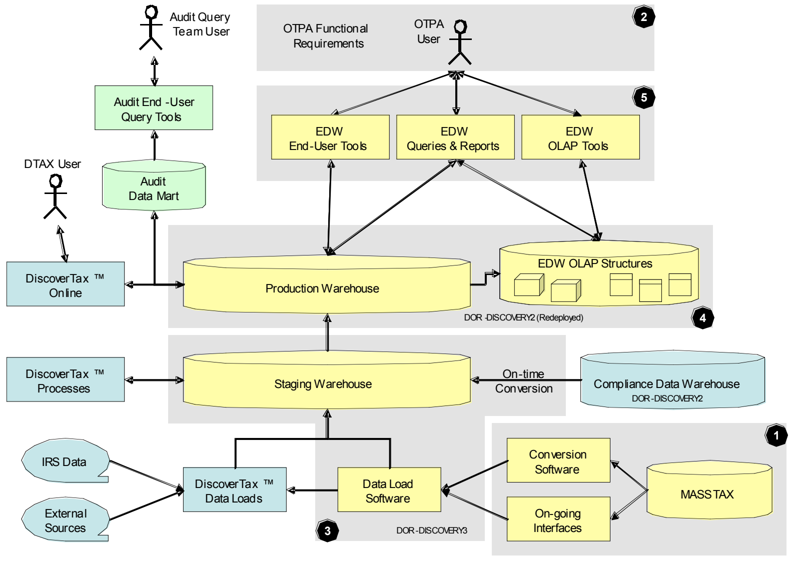
- Could simulate whole or a sub-set of return line to calculate alternative tax liability under various scenarios (exemptions change, deductions change, tax rate change, etc.)

- Work under way for improving its efficiency and usability.

Future

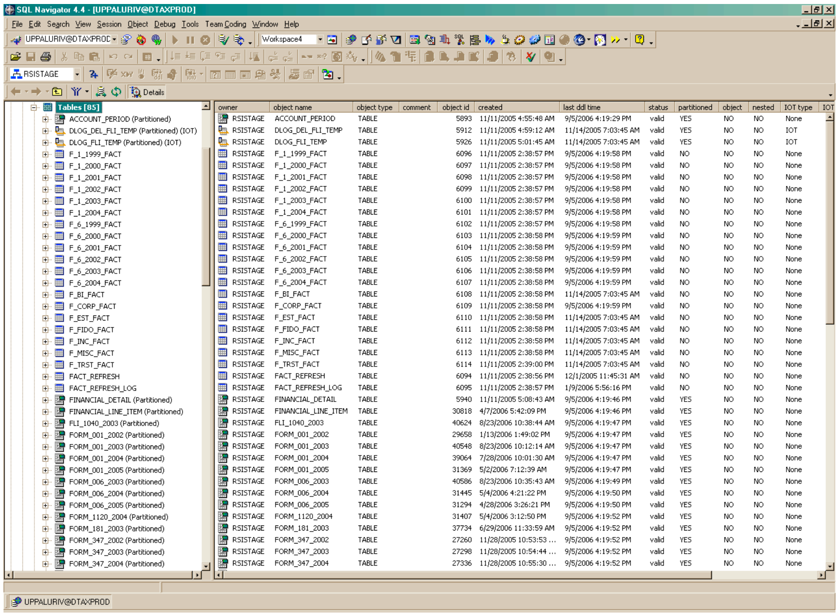
- Data warehouse is expanding (more data becoming available over-time). More storage space will likely be needed.
- Demand for data increasing by various divisions
- So, there are plans for investing more in the Data Warehouse
- Ultimate goal is to
  - Support a transition towards a more data-driven environment
  - Have even better performance measurement,
  - Have increased compliance management,
  - Improved and effective decision-making.

Flow Diagram



17

Warehouse Tables


SQL Navigator 4.4 - [UPPALURIV@DTAXPROD]
File Edit Search View Session Object Debug Tools Team Coding Window Help
UPPALURIV@DTAXPROD
Workspace4
RSISTAGE
Details
UPPALURIV@DTAXPROD

Tables [85]
- ACCOUNT_PERIOD (Partitioned)
- DLOG_DEL_FLI_TEMP (Partitioned) (IOT)
- DLOG_FLI_TEMP (Partitioned) (IOT)
- F_1_1999_FACT
- F_1_2000_FACT
- F_1_2001_FACT
- F_1_2002_FACT
- F_1_2003_FACT
- F_1_2004_FACT
- F_6_1999_FACT
- F_6_2000_FACT
- F_6_2001_FACT
- F_6_2002_FACT
- F_6_2003_FACT
- F_6_2004_FACT
- F_BI_FACT
- F_CORP_FACT
- F_EST_FACT
- F_FIDO_FACT
- F_INC_FACT
- F_MISC_FACT
- F_TRST_FACT
- FACT_REFRESH
- FACT_REFRESH_LOG
- FINANCIAL_DETAIL (Partitioned)
- FINANCIAL_LINE_ITEM (Partitioned)
- FLI_1040_2003 (Partitioned)
- FORM_001_2002 (Partitioned)
- FORM_001_2003 (Partitioned)
- FORM_001_2004 (Partitioned)
- FORM_001_2005 (Partitioned)
- FORM_006_2003 (Partitioned)
- FORM_006_2004 (Partitioned)
- FORM_006_2005 (Partitioned)
- FORM_1120_2004 (Partitioned)
- FORM_181_2003 (Partitioned)
- FORM_347_2002 (Partitioned)
- FORM_347_2003 (Partitioned)
- FORM_347_2004 (Partitioned)

owner	object name	object type	comment	object id	created	last ddl time	status	partitioned	object	nested	IOT type	IOT
RSISTAGE	ACCOUNT_PERIOD	TABLE		5893	11/11/2005 4:55:48 AM	9/5/2006 4:19:29 PM	valid	YES	NO	NO	None	
RSISTAGE	DLOG_DEL_FLI_TEMP	TABLE		5912	11/11/2005 4:59:12 AM	11/14/2005 7:03:45 AM	valid	YES	NO	NO	IOT	
RSISTAGE	DLOG_FLI_TEMP	TABLE		5926	11/11/2005 5:01:45 AM	11/14/2005 7:03:45 AM	valid	YES	NO	NO	IOT	
RSISTAGE	F_1_1999_FACT	TABLE		6096	11/11/2005 2:38:57 PM	9/5/2006 4:19:58 PM	valid	NO	NO	NO	None	
RSISTAGE	F_1_2000_FACT	TABLE		6097	11/11/2005 2:38:57 PM	9/5/2006 4:19:58 PM	valid	NO	NO	NO	None	
RSISTAGE	F_1_2001_FACT	TABLE		6098	11/11/2005 2:38:57 PM	9/5/2006 4:19:58 PM	valid	NO	NO	NO	None	
RSISTAGE	F_1_2002_FACT	TABLE		6099	11/11/2005 2:38:57 PM	9/5/2006 4:19:58 PM	valid	NO	NO	NO	None	
RSISTAGE	F_1_2003_FACT	TABLE		6100	11/11/2005 2:38:57 PM	9/5/2006 4:19:58 PM	valid	NO	NO	NO	None	
RSISTAGE	F_1_2004_FACT	TABLE		6101	11/11/2005 2:38:57 PM	9/5/2006 4:19:58 PM	valid	NO	NO	NO	None	
RSISTAGE	F_6_1999_FACT	TABLE		6102	11/11/2005 2:38:57 PM	9/5/2006 4:19:58 PM	valid	NO	NO	NO	None	
RSISTAGE	F_6_2000_FACT	TABLE		6103	11/11/2005 2:38:58 PM	9/5/2006 4:19:59 PM	valid	NO	NO	NO	None	
RSISTAGE	F_6_2001_FACT	TABLE		6104	11/11/2005 2:38:58 PM	9/5/2006 4:19:59 PM	valid	NO	NO	NO	None	
RSISTAGE	F_6_2002_FACT	TABLE		6105	11/11/2005 2:38:58 PM	9/5/2006 4:19:59 PM	valid	NO	NO	NO	None	
RSISTAGE	F_6_2003_FACT	TABLE		6106	11/11/2005 2:38:58 PM	9/5/2006 4:19:59 PM	valid	NO	NO	NO	None	
RSISTAGE	F_6_2004_FACT	TABLE		6107	11/11/2005 2:38:58 PM	9/5/2006 4:19:59 PM	valid	NO	NO	NO	None	
RSISTAGE	F_BI_FACT	TABLE		6108	11/11/2005 2:38:58 PM	11/14/2005 7:03:45 AM	valid	NO	NO	NO	None	
RSISTAGE	F_CORP_FACT	TABLE		6109	11/11/2005 2:38:58 PM	9/5/2006 4:19:59 PM	valid	NO	NO	NO	None	
RSISTAGE	F_EST_FACT	TABLE		6110	11/11/2005 2:38:58 PM	11/14/2005 7:03:45 AM	valid	NO	NO	NO	None	
RSISTAGE	F_FIDO_FACT	TABLE		6111	11/11/2005 2:38:58 PM	11/14/2005 7:03:45 AM	valid	NO	NO	NO	None	
RSISTAGE	F_INC_FACT	TABLE		6112	11/11/2005 2:38:58 PM	11/14/2005 7:03:45 AM	valid	NO	NO	NO	None	
RSISTAGE	F_MISC_FACT	TABLE		6113	11/11/2005 2:38:58 PM	11/14/2005 7:03:45 AM	valid	NO	NO	NO	None	
RSISTAGE	F_TRST_FACT	TABLE		6114	11/11/2005 2:39:00 PM	11/14/2005 7:03:45 AM	valid	NO	NO	NO	None	
RSISTAGE	FACT_REFRESH	TABLE		6094	11/11/2005 2:38:56 PM	12/1/2005 11:45:31 AM	valid	NO	NO	NO	None	
RSISTAGE	FACT_REFRESH_LOG	TABLE		6095	11/11/2005 2:38:57 PM	1/9/2006 5:56:16 PM	valid	NO	NO	NO	None	
RSISTAGE	FINANCIAL_DETAIL	TABLE		5940	11/11/2005 5:08:43 AM	9/5/2006 4:19:46 PM	valid	YES	NO	NO	None	
RSISTAGE	FINANCIAL_LINE_ITEM	TABLE		30818	4/7/2006 5:42:09 PM	9/5/2006 4:19:46 PM	valid	YES	NO	NO	None	
RSISTAGE	FLI_1040_2003	TABLE		40624	8/23/2006 10:38:44 AM	9/5/2006 4:19:47 PM	valid	YES	NO	NO	None	
RSISTAGE	FORM_001_2002	TABLE		29658	1/13/2006 1:49:02 PM	9/5/2006 4:19:47 PM	valid	YES	NO	NO	None	
RSISTAGE	FORM_001_2003	TABLE		40548	8/23/2006 10:12:14 AM	9/5/2006 4:19:47 PM	valid	YES	NO	NO	None	
RSISTAGE	FORM_001_2004	TABLE		39064	7/28/2006 10:01:30 AM	9/5/2006 4:19:47 PM	valid	YES	NO	NO	None	
RSISTAGE	FORM_001_2005	TABLE		31369	5/2/2006 7:12:39 AM	9/5/2006 4:19:47 PM	valid	YES	NO	NO	None	
RSISTAGE	FORM_006_2003	TABLE		40586	8/23/2006 10:35:43 AM	9/5/2006 4:19:49 PM	valid	YES	NO	NO	None	
RSISTAGE	FORM_006_2004	TABLE		31445	5/4/2006 4:21:22 PM	9/5/2006 4:19:50 PM	valid	YES	NO	NO	None	
RSISTAGE	FORM_006_2005	TABLE		31294	4/28/2006 3:26:21 PM	9/5/2006 4:19:50 PM	valid	YES	NO	NO	None	
RSISTAGE	FORM_1120_2004	TABLE		31407	5/4/2006 3:12:50 PM	9/5/2006 4:19:52 PM	valid	YES	NO	NO	None	
RSISTAGE	FORM_181_2003	TABLE		37734	6/29/2006 11:33:59 AM	9/5/2006 4:19:52 PM	valid	YES	NO	NO	None	
RSISTAGE	FORM_347_2002	TABLE		27260	11/28/2005 10:53:53 ...	9/5/2006 4:19:52 PM	valid	YES	NO	NO	None	
RSISTAGE	FORM_347_2003	TABLE		27298	11/28/2005 10:54:44 ...	9/5/2006 4:19:52 PM	valid	YES	NO	NO	None	
RSISTAGE	FORM_347_2004	TABLE		27336	11/28/2005 10:55:30 ...	9/5/2006 4:19:52 PM	valid	YES	NO	NO	None	