出國報告（出國類別：考察）

環境資料管理及資料庫設計訓練

服務機關：行政院環境保護署

姓名職稱：何璠　　監資處技士

派赴國家：美國

出國期間：民國 94 年 7 月 4 日至 7 月 17 日

報告日期：民國 94 年 10 月 5 日

行政院及所屬各機關出國報告書提要

出國報告名稱：環境資料管理及資料庫設計訓練

頁數：23 頁（未含附件）

出國計畫主辦機關/聯絡人/電話：

經濟部國際合作處/羅研究員特/02-23513855

出國人員姓名/服務機關/單位/職稱/電話：

何璠/行政院環境保護署/環境監測及資訊處/技士/02-23117722-2329

出國類別：■1 考察□2 進修□3 研究□4 實習□5 其他

出國期間：民國 94 年 7 月 4 日至 7 月 17 日　　出國地區：美國

報告日期：民國 94 年 10 月 5 日

分類號/目：

關鍵詞：環保事實資料倉儲、環境資料庫系統

內容摘要：本次出國計畫共前往三處，其中兩處位於華府的美國環保署總部之環保資料系統中心以及環境資料室，另一處則位於北卡羅來納州研究三角公園(Research Triangle Park)的國家電腦中心，此行目的是蒐集美國環保署之環境資料管理及資料庫相關技術，包括環保事實資料倉儲技術、環境地理資料庫建置、環保資料庫系統、環境資料交換與分享、伺服器管理等主題，藉以瞭解美國環境資料管理的發展與技術，作為本署評估未來環境資料庫技術之發展方向。

目錄

第一章 前言

1. 緣起

自環保署成立以來，為配合環境品質管理施政之需求與跨部會國土資訊系統之建置，逐步規劃並陸續建置各類環境品質資料庫與環境資訊系統，包括空氣、水、廢棄物、毒化物、環境品質、環保統計等相關系統。民國 89 年年底開始希望規劃一個健全的環境資料庫系統架構與建置體系，期能據以建置一個能夠即時提供完整、正確且有效資料與資訊的環境資料庫。環境資料庫體系建置工作並非要廢棄現有既存的資料庫及系統而重新建立新的資料庫與系統，而是整合各單位已建置之環境資料，使得民眾、跨業務處以及政府跨部會業務所需之資料有一個共通的提供管道。

建置環境資料庫是一個跨部會、跨領域、且需要長時間的工作，美國環保署自 1993 年開始整合，至今十餘年已小有成果；我國之環境資料庫建置工作自民國 90 年開始規劃，91 年開始著手整合，目前初步建置之重點有二，第一類為環境品質類，包括空氣品質資料、河川水質資料、水庫水質資料、垃圾清運及回收量、以及自來水水質合格率資料；第二類為被環保署列管、具管制編號之點污染源類資料，包括空、水污染源、廢棄物產生源、以及毒化物運作源資料。整合後之資料將依不同使用對象，分階段逐步公開資料提供環保管制、以及各界應用參考，目前建置的這些資料說明如下：

一、點污染源資料

點污染源環境資料是指被環保署列管、具管制編號之點污染源類資料，這些資料均由一特殊值串連，該值稱為『管制編號』，其使得資料間之關係得以明確定義，尤其是與這些點污染源之來源資料庫間之資料對應關係，扮演著相當重要的角色。點污染源資料所建置內容分述如下。

(一)基本資料

此部分資料集合了環保署各業務單位管轄之場(廠)所及其基本資料，記錄對象包括空、水污染源、廢棄物產生源、以及毒化物運作源資料；記錄內容包括場(廠)所之名稱、所在位置、所屬行業、廢水排放承受水體、電話、E_MAIL、負責人資料、證號資料、座標資料以及各業務單位列管狀態等內容。

(二)稽查及處分資料

此部分集合了環保機關對管轄之場(廠)所，所進行空、水、廢之稽查作業所產出之稽查工作單資料，以及環保機關依法執行開立處分之處分書資料，記錄內容包括稽查處分之日期、項目、違法內容、以及後續處理等內容。

(三)水污染資料

此以廢(污)水污染特性為資料建置標的，記錄內容包括廢(污)水排放許可證核發狀態、許可廢(污)水排放水量、許可處理污泥量、各排放口水質、以及各場(廠)所廢(污)水防治設備數量資料等內容。

(四)空氣污染資料

此以空氣污染特性為資料建置標的，資料來源為環保署空保處，記錄內容包括空氣許可證核發狀態、空氣污染整廠排放量、空氣污染煙囪及逸散排放量、以及各場(廠)所空氣污染防治設備數量資料等內容。

(五)廢棄物資料

此以廢棄物為資料建置標的，記錄內容包括各場(廠)所產生之廢棄物類別及數量資料、以及各場(廠)所暫存於場(廠) 內之廢棄物類別及數量資料現況等內容。

(六)毒化物運作資料

此以毒化物資料為建置標的，記錄內容包括毒化物運作許可證核發狀態、毒化物操作量以及毒化物釋放量資料等內容。提供各類毒化物釋

放於環境之各類方式以及其量化資料，包括生物濃縮性毒化物、致癌性毒化物、暴露性危害性毒化物、非立即危害性毒化物之毒化物量。

2. 目的

環境資料庫是以民眾、跨業務處之需求為建置主軸，彙整各處之重要資料，為資料庫的長遠建置作業立下基礎，環保署實體整合之環境資料於 91 年度初步落實建置，開始匯入資料，並建立來源資料與環境資料庫間之傳輸通路，並開發資料供應系統，提供原始資料查詢、統計繪圖、以及繪製位置圖等功能。整合後之資料將依不同使用對象，分階段逐步公開資料提供環保管制、以及各界應用參考。因此此行希望能蒐集美國環保署在環保資料庫建置的相關經驗，包括環境資料交換、GIS 地理資料庫等相關資訊，以供我國在建置相關系統時的參考資料。

3. 考察行程與出國人員

本次出國參訪自 94 年 7 月 4 日～7 月 17 日，出國期間參訪單位包括位於華府的美國環保署環保資料系統中心、環境資料室，以及位於北卡羅來納州研究三角園區(Research Triangle Park)的國家電腦中心，以下是本次參訪單位人員簡要說明以及詳細參訪行程：

姓名	單位/主管業務或職稱
Mr. Steve Hufford	美國環保署環保資料系統中心/系統工程計畫經理
Mr. Richard A. Martin	國家電腦中心/副主任
Mr. Robin L. Gonzalez	環境資料室/主任
Mr. Pat Garvey	美國環保署/資訊計畫經理
Mr. Dave Wolf	美國環保署/資訊計畫經理

表一 參訪行程

訓練進修日期及時間 (Visiting Time)	訓練進修地點 (Location)	訓練進修目的及討論主題 （Topics for Discussion）
7/4 2005	Taipei – Washington DC 台北－華府	去程
7/5-7/10, 2005	Washington DC 華府	1. Learn how to design Toxics Release Inventory Database for users to easily complete TRI forms. 學習如何設計毒物排放庫存資料庫，以供使用者能夠順利完成 TRI 格式之填寫 2. Learn how to develop and manage the Envirofacts Data Warehouse. 學習如何設計並管理環保事實資料倉儲 3. Learn how to build the environmental database system. 學習如何建構環境資料庫系統 4. Learn how to build database with Geographical Information Support 學習如何在 GIS 的支援下建構資料庫
7/11	Washington DC - Durham, NC 華府-德漢，北卡羅萊那州	前往北卡
7/11-7/12, 2005	Durham, NC 德漢，北卡羅萊那州	1. Visit National Computer Center to exchange experience in database management 參訪美國國家電腦資訊中心並交換資料庫管理經驗 2. Collect information about server management 蒐集伺服器管理之相關資訊
7/12 2005	Durham, NC- Washington DC 德漢，北卡羅萊那州-華府	前往華府

7/13-7/14, 2005	Washington DC 華府	1. Learn how to have Information quality assurance in the database 學習如何確保資料庫內資料之品質 2. Learn how to collect environmental information and share the database to other organization 學習如何蒐集環境資料至資料庫，並將資料庫分享給其他機構
7/15-7/17	Washington DC -Taipei 華府-台北	返程

第二章 美國環保署環保資料系統中心考察紀要

1. 國家環保資料系統中心簡介

位於華府的美國國家環保資料系統中心為美國環保署在資訊工程方面最大型的系統中心，其又與以下四處衛星站相連結：

- 研究三角園區（Research Triangle Park, NC）
- Richland, WA
- EStennis 太空中心(EStennis Space Center, MS)
- 拉斯維加斯(Las Vegas, NV)

除此之外，另外還連結至其他 10 個區域，以確保現有及新的使用者能便利取得相關服務。下圖是整個中心於全美的分佈圖。

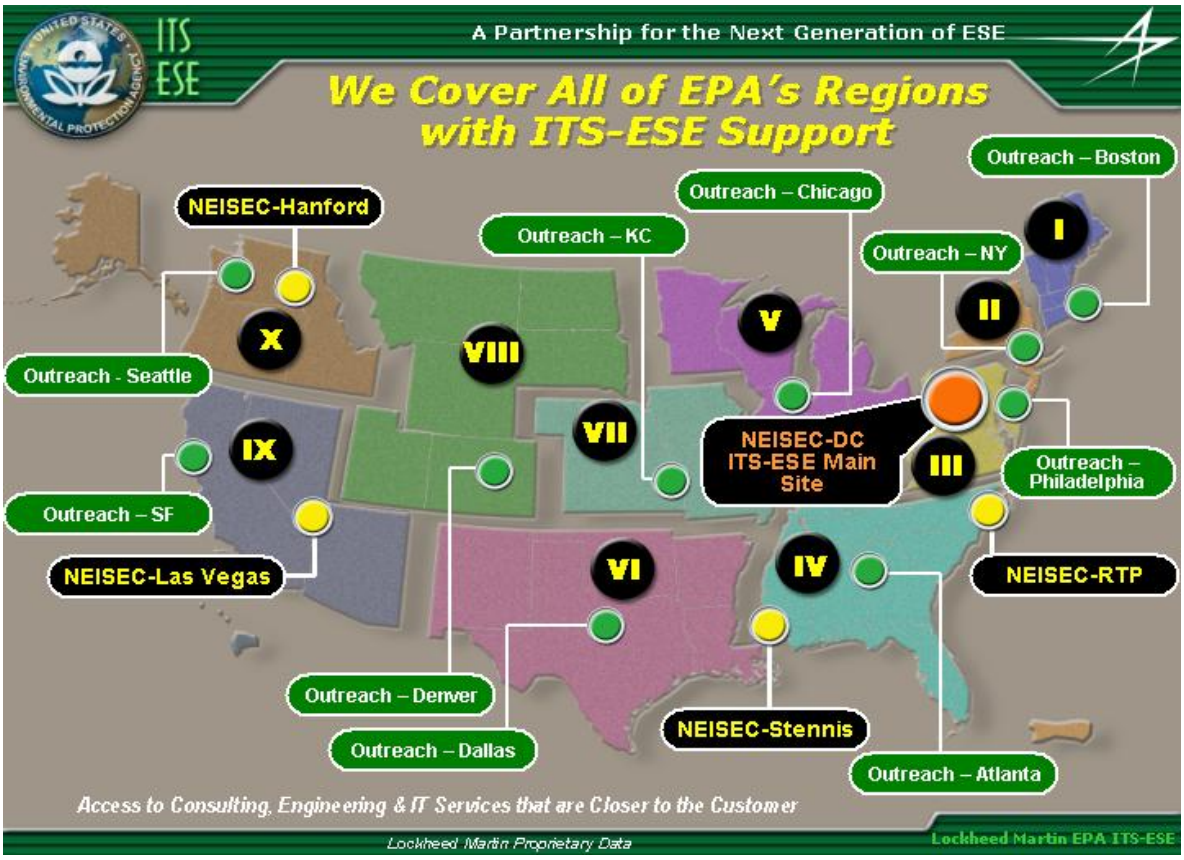


圖 2-1、美國國家環保資料系統中心之分支於全美分佈圖

圖 2-2、美國環保署總部

2.參訪蒐集資料摘要

此次參訪此中心，主要蒐集該中心正在執行的四項計畫：1.毒物排放資料庫。2.環保事實資料倉儲。3.環境資料庫系統。4.GIS 資料庫。

(1)毒物排放資料庫（Toxic Release Inventory）

毒物排放資料庫（簡稱 TRI）計畫的形成主要是基於社區居民對於附近工廠所排放危害性化學物質資訊有取得之需求而來。針對此目的，於 1986 年時又制訂了緊急事件規劃與社區有知道權力法案（Emergency Planning and Community Right-to-Know Act，簡稱 EPCRA)。此法案的 311,312 章節要求工廠要向州及地方政府申報公司地點和化學物貯存量等資訊，使社區居民在將來發生意外或化學物溢散時能提早準備防範。第 313 章節則要求環保署和州政府每年蒐集化學物排放和運輸量的資訊，同時將此資訊公開。1990 年國會通過另一項法案，污染預防法案（Pollution Prevention Act)，此方案要求在既定的 TRI 計畫下，需另外再提供廢棄物管理和污染源減量的資訊。美國環保署除了蒐集數據

到 TRI 外，還需將資料編纂至一個大家都可以使用的介面工具上，例如 TRI 瀏覽器或環境事實資料庫。最近 EPA 又訂定一些新的規則，粗估可將涵蓋的化學物種類增加到 650 類左右，另外又增加 7 種新的行業，同時一些具有持久性、生物累積性、毒性化學物（如 PBT）的申報門檻也被降低，如此可使民眾取得這些化學物的資訊。

為了讓業者確實瞭解 TRI 如何申報，除了在各地舉辦訓練講習會外，EPA 特別為需要申報業者開發了一套軟體，稱為 TRI-ME，申報者可自行在網路上下載，並依據指示申報，跟我國的所得稅申報軟體頗為類似，主要目的在於減少人工申報容易產生的錯誤。此軟體具備檢查錯誤的功能，申報者填寫完之後，可將結果列印出來或以儲存至磁片郵寄的方式，另亦可透過電子資料傳輸方式透過中央資料交換系統（Central Data Exchange）傳送到 EPA。

有關使用者查詢部分，EPA 開發一套以 Web 介面的查詢平台，稱為 Tri-Explorer。此套資料查詢系統非常容易使用，使用者可取得多面向的資訊，只要用滑鼠輕輕按個兩三次就可以產生一個實用的報表，不像其他應用程式的學習曲線那麼長。

Tri-Explorer 的網路服務（Web Service）建置是採用 SAS/IntrNet 軟體，其包含之三項組成如下：使用者介面（通常是網路瀏覽器）、網路伺服器、後端資料庫伺服器（Back End Database Server）。

由於環保相關的資料實在太多，在各個資料庫之間若能利用其關連性則可減少資料重覆鍵入所耗費的成本，例如在 TRI-Explorer 內使用了環保事實資料庫（Envirofact）中的一些表格，而環保事實資料庫亦使用了 TRI-Explorer 內的表格，另外，在使用 TRI-Explorer 時，亦可擷取環保事實資料庫中的資料。

有關 TRI-Explorer 的網路服務(Web Service)，其應用程式的派遣方式(Application Dispatcher)詳如下圖：

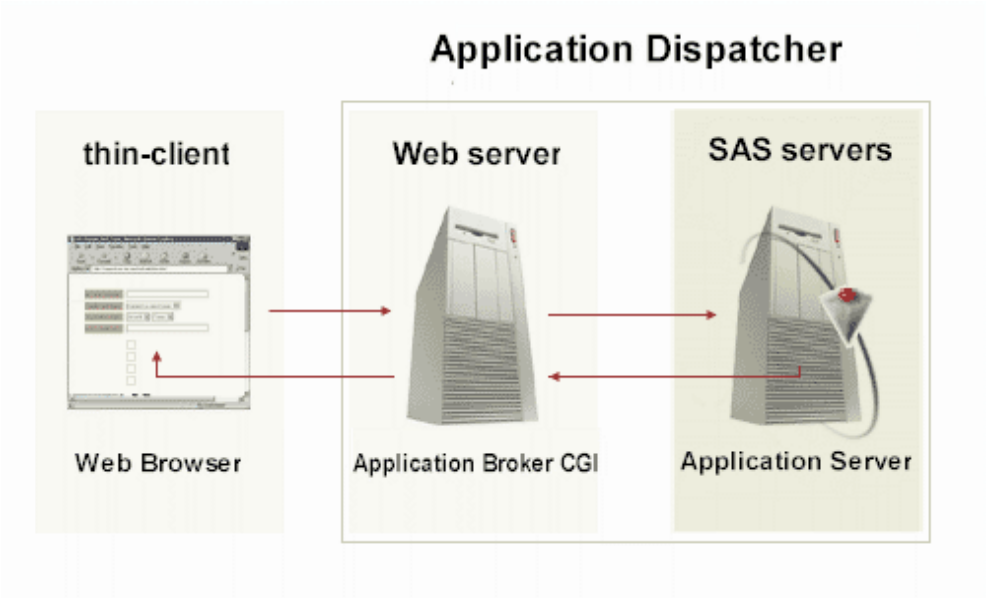


圖 2-3、TRI-Explorer 其應用程式的派遣方式(Application Dispatcher)

(2)環保事實資料倉儲(Envirofact)

環保事實資料庫是一個 EPA 環保資料查詢系統，開放給大眾查詢相關環保資訊，包括空、水、廢、毒、輻射、土壤、地圖等資訊。

其實環保事實資料倉儲計畫的由來，主要基於以下四點考量：

a. 跨計畫之間的整合

b. 污染預防

c. 地圖資料的倡導

d. 自願性遵守法規

e. 以社區為基礎的環保觀念

其中最重要的就是各個計畫的整合，整合工作做的越徹底，則愈容易達成其他目標，一般來說，環保事實資料倉儲是由下圖中各方塊中的資料庫組合而成：

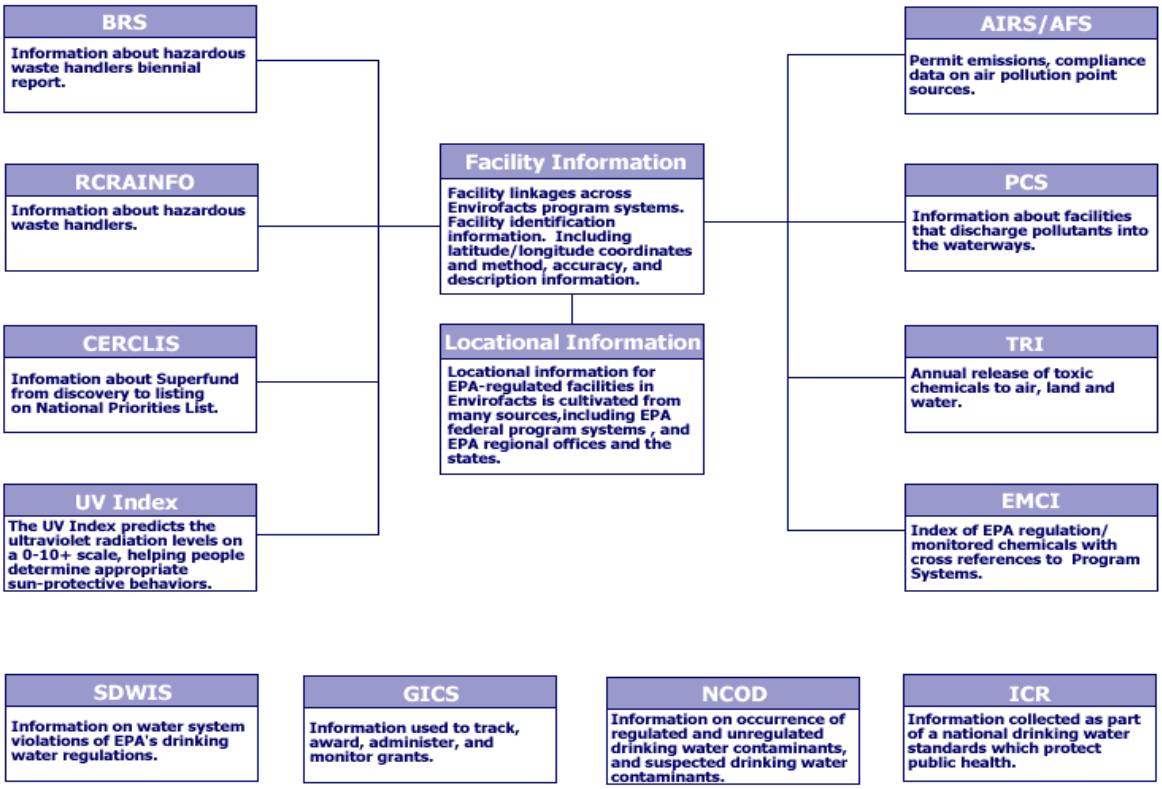


圖 2-4、環保事實資料庫之資料模式

(3)環境資料庫系統(Window to my Environment，簡稱 WME)

美國環保署發展的 WME 與本署現著手建置的環境資料庫頗為類似，其主要使用者乃設定於社區居民，因為他們對周遭環境潛在污染源有知的權利。由於此資料庫內資料深入地方社區，因此這套系統是在聯邦政府、州政府以及地方組織的通力合作下完成，資料共涵蓋了美國環保署、州政府、社區、受管制社區、以及地圖系統等資訊。如此將豐富的環保資訊整合在詳細的區域地圖將有助於發掘潛在的環境污染，並提早防制以維護社區居民的健康。

此環境資料庫系統主要功能如下：

a. 最新的互動式地圖：此套 WME 可依使用者的喜好設定地圖標示出管制設施、監測站、水體、水流、地下水、人口密集區的位置，當然

還有傳統地圖中所具備的，例如街道、高速公路、學校、教堂的位置。瀏覽功能可用放大、縮小或平移來控制。

b. 含括周遭（Ambient）環境狀況之資訊：此系統提供了查詢區域的紫外線數值，以及防曬的建議，亦從附近的空氣或水質監測站擷取監測資料供使用者參考。另亦提供該區域的表土特性。

c. 具備分析與報表輸出的功能：此系統可連接至美國環保署的環境事實資料庫、TRI-Explorer 以及”在你的河川沖浪”(Surf Tour Watershed)的工具，還有各州使用的相關工具，例如賓州的”E 事實”(E-Facts)和德拉瓦州的”環境領航員”(Environmental Navigator)，這些工具都可以讓使用者自行量身製作所需資訊的報表。

d. 地方政府的服務與聯絡：此系統可連結至官方與非官方組織，讓使用者可以聯絡與查詢當地相關資訊。

(4) GIS 資料庫（Geographic Information Systems）

以目前的資訊科技來說，資料庫最終還是要與 GIS 結合，就如上述的資料庫系統一樣，美國環保署亦有一獨立單位負責規劃並定期更新全美的 GIS 資料庫，由於此資料庫所需資料需相當精細且需及時更新，此單位持續維持與相關部門（例如美國國防部）的協調與合作。目前我國內政部亦著手建立國土資訊系統，此亦為一 GIS 資料庫，因此應可參考美國的經驗。

所謂 GIS，它是一種可以用來分析和通覽地理地形的技術。GIS 可將地點與資料和圖層相連接，讓使用者可清楚瞭解當地的地理狀況，當然亦可依據需求選擇不同的圖層。

GIS 通常會讓人聯想到地圖，然而地圖僅是它三項功能中其中一項。三項功能觀點詳述如下：

a. 資料庫觀點：GIS 是一種非常獨特的資料庫，可稱為地理資料庫（簡稱 geodatabase）。基本上來說，GIS 乃建構於結構化的資料庫。

b. 地圖觀點：GIS 是一組地圖與展望圖，有展現地貌的功能。它可支援查詢、分析與編輯資訊的功能。又可稱為地理形象化 (geovisualization)。

c. 模式觀點：GIS 是一組資訊轉換工具，從現有的資料轉換而得到地理資料。此轉換過程稱為地理處置(geoprocessing)，從現有資料庫取得資料後，利用分析的功能，最後將獲得的結果寫入新的資料庫。

目前美國環保署利用 GIS 技術，將不同的圖層套用在全美地圖上，用途包括了 9 大項：空、水、土壤、沈降、污染物排放、氣候、敏感資源、人口、區域邊界。

第三章 美國國家電腦資訊中心考察紀要

1. 美國國家電腦資訊中心簡介

圖 3-1、美國國家電腦資訊中心

美國國家電腦資訊中心位於北卡羅萊那州的研究三角園區（Research Triangle Park, NC）。研究三角園區為美國 IT 產業研發重鎮，許多知名公司在此進行科技研發，此園區是全球最大的模範綠建築(green buliding)，曾獲得多項獎項，因其擁有數百項環境友善的功能。茲列舉其中數項綠色環保特色：

a. 對於園區通勤員工，提供誘因使大家儘量使用共乘制度，減少了 25%的停車面積。
b. 較一般建築物能源使用效率高 40%。
c. 減少 50%道路清掃及設施清潔量。
d. 它的照明系統較一般標準系統效率高 70%。
e. 80%營建廢棄物被回收再利用。
f. 90%的道路照明能源來自於太陽能。
g. 100%收集並處理暴雨逕流。

h. 園區中央空調100%不使用造成臭氧層耗竭的冷媒。

i. 建築物使用週期增加200%(一般使用週期為30到50年,但園區建築物則可使用100年)。

美國國家電腦資訊中心內擁有全美最大的電腦機房,全美大部分重要的資料庫都被傳送到此處進行備援。

2.參訪蒐集資料摘要

本次參訪北卡美國環保署國家計算機中心最主要目的在於了解美國環保署電腦機房之實體安全管理措施及資訊安全之管理技術。

美國經911恐怖攻擊事件後為防範恐怖份子再度攻擊事件,赴美前至美國在台協會辦理美國簽證就感受到安全檢查之嚴格,國際線飛機到美國及美國國內線飛機之轉機安全檢查更是嚴格,小刀等危禁品均予以沒收就連身上或鞋子有金屬東西均需拿下經X光機檢查,故機場通關花費大量人力及時間,可說是相當不便。

到了北卡美國環保署國家計算機中心本部,大樓進出也有金屬探測器進行檢查,訪客需內部人員帶領才能進入,國家算機中心機房大樓進入之訪客亦需經櫃檯人員以金屬探測器一一檢查,進出除登記外,照像機亦需拿出由櫃檯統一保管,大樓內不准拍照。檢查過程相當複雜耗時,但本次到美國環保署國家計算機中心主要目的之一即是觀摩美國環保署機房之實體安全管理,所以如考量實體安全方面,此做法應屬比較嚴謹之方式,故其取捨可作為本署機房門禁管制規劃之參考。

美國環保署國家計算機中心負責管理及操作環保署各單位所開發之資訊系統,經費由各單位編列,依所需提供之服務項目將經費撥給環保署國家計算機中心,由該中心負責電腦機房軟硬體設備之建置及操作維護。因美國幅員遼闊故其機房之規模與本署相比可說相當龐大,且資訊安全防護於現行以網路為主之資訊系統來說可說是相當重要,美國環

保署國家計算機中心為維護電腦機房之資訊安全，故電腦機房中系統之操作管理及資訊安全方面均委外或進駐相當多的人力以進行資料備份、程式更新等例行之操作及管理作業，資訊安全維護方面，日常之網路監控委由外包公司透過網路進行 24 小時之資安監控，因資安事件日新月異，防範又需投入相當之人力，故委外進行資安防護監控將是未來確保資訊安全可行之方式。

美方委外合約執行年限較長可由 3 年延長至 9 年，故其通常花費較長之時間(約 1~2 年)進行合約內容之規劃及處理合約爭議訴訟事宜，委外合約之規劃先進行完整考量，可以避免執行時發生意外狀況，花費額外之處理成本，亦值得我們執行委外合約發包之參考。

本次參訪時並由美方介紹該中心資料安全防護的策略，在安全防護管理方面分為資訊安全管理、資料存取控制、網路管理、主機管理及資料管理五個階層。

資訊安全管理方面，因美國幅員遼闊故其機房之規模與本署相比可說相當龐大，且資訊安全防護於現行以網路為主之資訊系統來說可說是相當重要，美方除委外進行 24*7 全天候之資安監控以提找發現外部之攻擊行為及安全之漏洞進行防護，也由計算機中心內部進行資安監控並定期檢視各主機之日誌檔，以發掘可能之防護漏洞。除此之外也人員安全管理、網路伺服器安全風險估計、意外事故應變計畫及災難復原均有事先嚴格之規範與計畫。

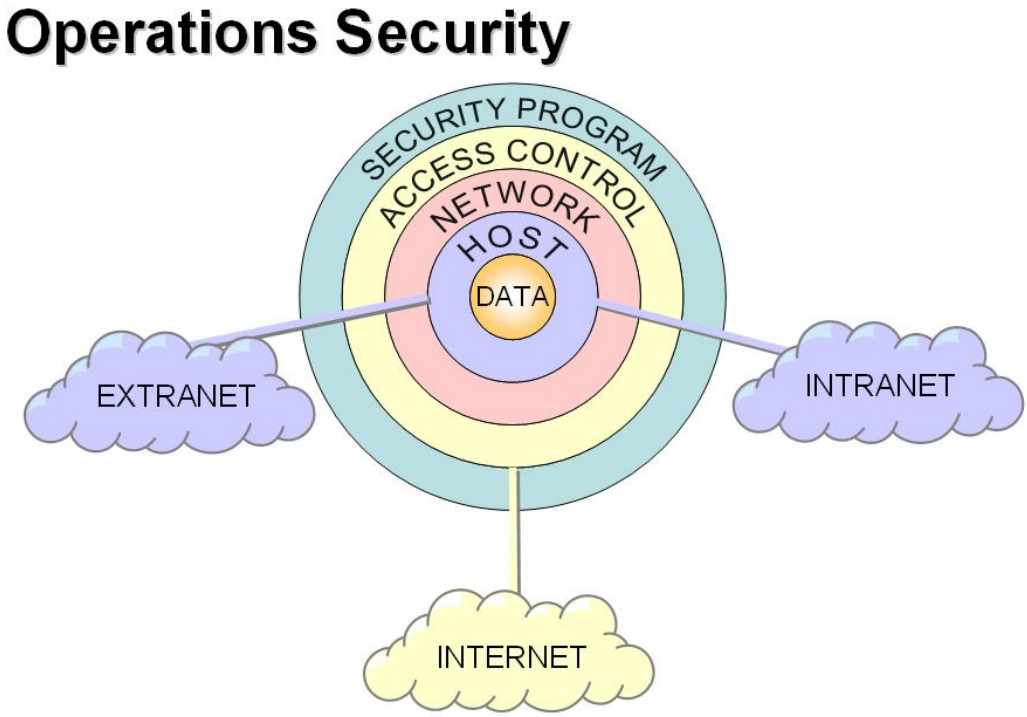
資料存取控制方面，以具備帳號密碼及時序號碼產生器二種機制的嚴謹之身分認證授權，有些縱要資料更需限制存取之地點，對所有資料均需明確標示並訂定存取之管控機制。

網路管理方面設置有路由器、防火牆、入侵偵測等設備以阻擋不被許可之設備存取及防止過濾外部之攻擊行為，對重要資料網路傳輸也要求需加密，以防止遭網路竊聽，資料被竊取。

主機管理方面主要做法有標準化環境規劃統一主機之設定環境、安裝防毒軟體進行主機病毒之防護、定期進行主機安全性之掃描、對特權使用者帳號管控並記錄其存取之內容以供查詢、對主機安全政策變更均需留下紀錄以便日後追查及確保主機之安全。

資料安全管理方面重要資料以加密儲存、資料儲存均有完整標示及分割儲存以避免資料不被竊取。

資料安全防護是全面性的，一環節有安全之漏洞均有可能造成資料之外洩，下圖為其資料安全防護策略之架構圖。



於北卡美國環保署國家計算機中心另有一小組進行環境視覺模擬，本署參訪時美方人員展示以 3D 數值模型來模擬 911 恐怖攻擊事件雙子星大樓倒塌污染擴散情形，系統以流體力學之模型來計算污染源之擴散情形，以電腦模擬模式計算後再以電腦 3D 動畫方式展現，以視覺模擬方式來展現具相當真實感，此技術將可運用於空氣污染或毒化物擴散等之模擬與預警等方面。

第四章 美國環保署環境資料室考察紀要

1.環境資料室（Office of Environmental Information）簡介

圖 4-1、與接待人員合影於環境資料室

環境資料室位於華府美國環保署總部，該單位的任務就是確認所蒐集的環保資料具有高品質，並將其對大眾公布。為了達到這個目標，該單位協助制訂相關指引(Guidance)，包括蒐集、管理、分析及提供環保資料的方式。此次前往環境資料室，主要是考察其環境資料蒐集的部分。

2. 參訪蒐集資料摘要

本次參訪此單位的目的，在於蒐集有關資料品質的確保，以及資料蒐集時的傳輸流程與分享機制。

(1)資料品質確認

對於資料的處理，以及確認其品質，需先制訂資料的標準，對此 USEPA 制訂了一系列的資料註冊系統(DATA Registry)，這些註冊系統的角色如下：

a. 能清晰地描述、貯存、分析及分類資料。

b. 支援資料標準發展的程序與分布。

c. 提供良好型態的資料元件與價值領域（Value Domain）的來源。

d. 促進署內系統資料的再使用性。

e. 促使資料可分享、整合與相互比較。

f. 作為署內資料的資料屬性(metadata)的核心來源。

以下介紹規範資料標準的六項註冊系統：

a.應用程式與資料庫註冊系統

此為資訊系統的庫存所在，並在此處整合所有的註冊系統，內含有資訊來源的相關目錄。

b.專有名詞參考系統(Terminology Reference System)

內含美國環保署資料庫內所有字彙的定義，關鍵字及其構想。字彙所涵蓋的範圍不僅包括了多語言環境總詞彙(General Multilingual Environmental)外，亦包括了署計畫辦公室、資訊系統、法規、州蒐集資料處的字彙。總計字彙來源達 229 處，共 11977 個專有名詞。

c.XML 註冊系統

美國環保署與交換網路中心(Exchange Network)目前已建置臨時性 XML 註冊系統，此系統儲存了交易合約內容，也就是環保署與各州之間的正式合約內容。當然此系統亦包含了 XML 綱要(Schema)和支援的資料屬性。

d.環境資訊管理系統

此系統包含了一般資料外，亦包含了地理資訊以及 ORD 資訊產品的相關資訊。並依據 ISO/IEC 19115 規定建置。

e.設施註冊系統(Facility Registry System)

此系統可查詢設施的地點，主要依據設施身份確認標準來建置。

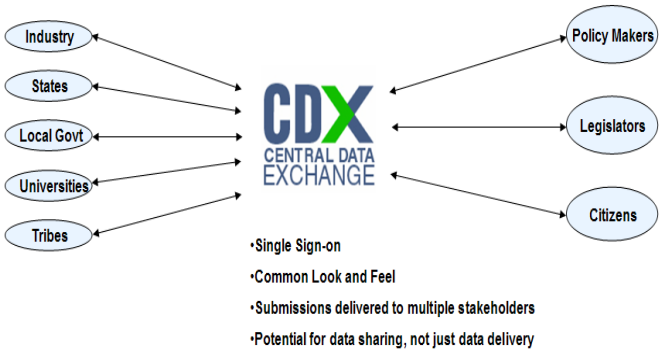
f.物質註冊系統(Substance Registry System)

此系統包含了化學品相關資訊，與我國的物質安全資料庫系統相當類似。除此之外，亦包括了生物有機體的資訊，以及與法規有關的化學

品資訊。

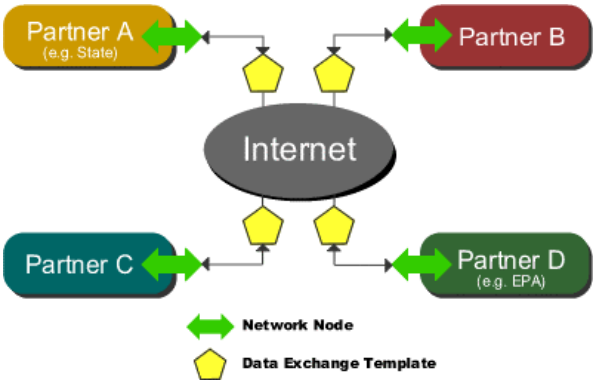
(2)資料傳輸與分享

美國環保署對於環境資料的傳輸與分享是靠環境資訊交換中心(CDX, Central Data Exchange)，它是運用網際網路(internet)及 21 世紀的最新資料技術與工具(架構示意圖如右)，藉以提升電子資料交換的安全。目前已超過 3 萬個合作夥伴在 CDX 註冊，其中今年註冊的就超過 8 千個。全美國已有 49 個州運用 CDX 所提供的空氣、水…等相關環境資料。



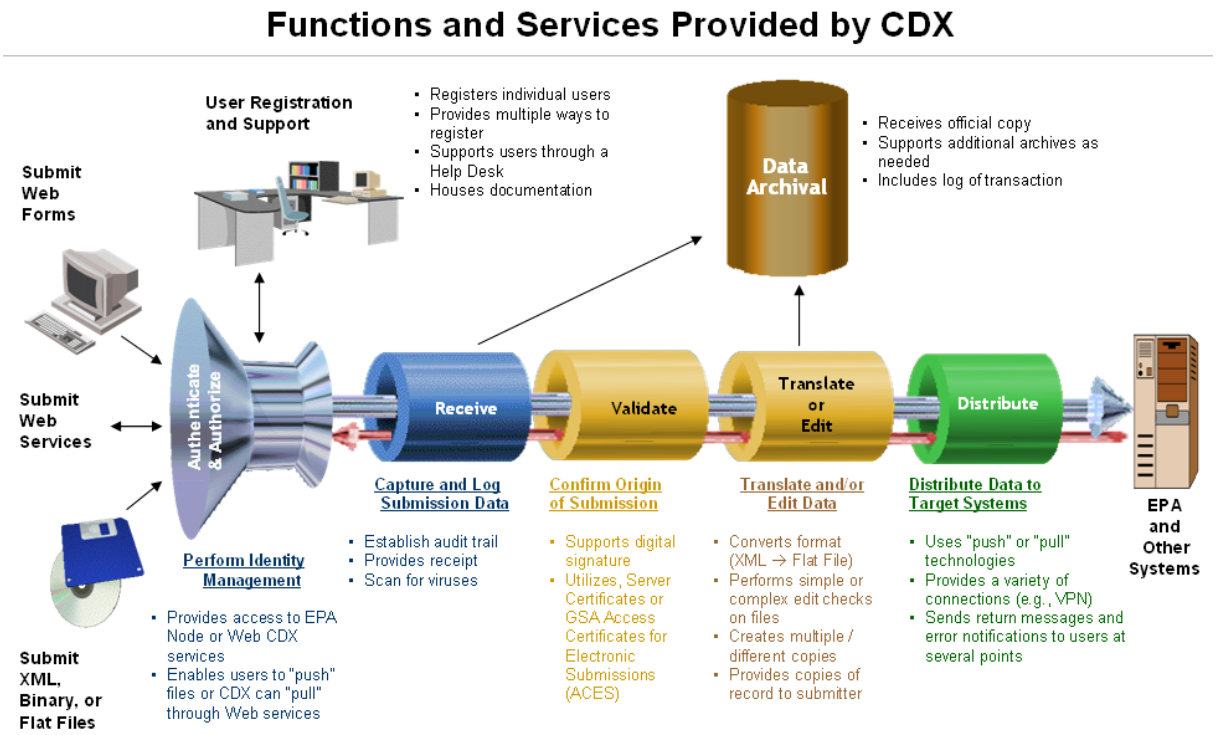
CDX removed the complexities and standardized data exchanges

在此機制下，並非所有的資料均集中存放在 CDX 中心，而是分散儲存於各相關單位，然後透過網際網路(internet)之 XML Web servers 方式，即時讀取相關資料後傳送至使用者手上(如圖左)。也因為相關資料均即時至各主管機關抓取，所以資料均可維持在最新狀態。



環境資訊交換中心的目標是建立快速、有效率、安全且精確的即時資料交換網絡，藉以連結聯邦政府、地方政府、工廠及美國環保署的各個分支單位。由於其成效良好，CDX 已為這些相關單位或組織節省了許多處理資料的時間與金錢。此系統也於 2004 年十月獲得「政府電腦新聞雜誌」(GCN, Government Computer News)之年度便民服務大獎、傑出政府部門資訊技術獎(Excellence in Government Information Technology Award)以及 2005 年年度優良技術大獎(Award for Technical Excellence)。

就技術面來說，CDX 包含了二種不同的傳輸技術，一般使用者可以輕鬆地透過網際網路去傳送及接收環境資料與相關報表，而另一方面，CDX 也允許系統對系統(machine-to-machine)的資料傳輸方式，以符合不同的使用者需求。



CDX 的核心技術摘要：

- Java 2 Enterprise Edition Platform
- Advanced Web Services with direct internet message encapsulation(DIME)
- XML Key Management Services (XKMS)
- Open Source Software

儘管 CDX 使用了美國環保署最優秀且先進的技術，但是為了兼顧各種不同的使用者需求，他們仍然必須花費龐大的人力進行離線資料處理，在我們參觀 CDX 的最後行程中，發現他們雇用了大批高中程度的人員，每天處理上百箱的郵寄信件，這些信件是索取資料用的，雖然 CDX 的網路功能已非常完善，但他們預估短期內這種人工作業尚無法完全由電腦網路取代。

第五章 考察心得

美國歷經 911 恐怖攻擊事件後，為防範恐怖份子再度攻擊，相關安檢作業變得非常嚴格，不論是遠在台灣的美國在台協會或美國的機場及相關政府單位，只要是進入室內場所均需通過嚴密的身分確認與安全檢查，甚至部分單位還需通過金屬探測器或 X 光機檢查，不僅美國方面需投入大量人力及物力，對於訪客來說也是相當不便。恰好此次參訪期間又碰巧遇到恐怖份子攻擊倫敦地鐵，造成重大傷亡事件，以美國來說對此恐怖攻擊更是繃緊神經，安檢的人力部署更是一下子就增加了數倍，讓我大開眼界。雖然對此需要花費相當多的經費從各方面來防止恐怖攻擊，且對海關及旅客通關會增加許多不便，但至少迄今可證明這些措施是有效果的。

此次行程為參訪性質，但受訪單位對此次的會面頗為重視，並希望進一步與環保署進行技術交流與合作。美方做事的敬業態度亦令我印象深刻，對於提出的技術相關問題有問必答，且提供相當豐富的簡報資料，相信這對於台美合作已踏出了第一步。

附錄一

Capability Maturity Model – Integrated Strategy

附錄二

Facility Registry System

附錄三

Contracting for IT Services

附錄四

Toxic Release Inventory

附錄五

Envirofacts

附錄六

Data Marts/Data Warehousing

A Partnership for the Next Generation of ESE

"Consistently Exceeding the Customer's Expectations"

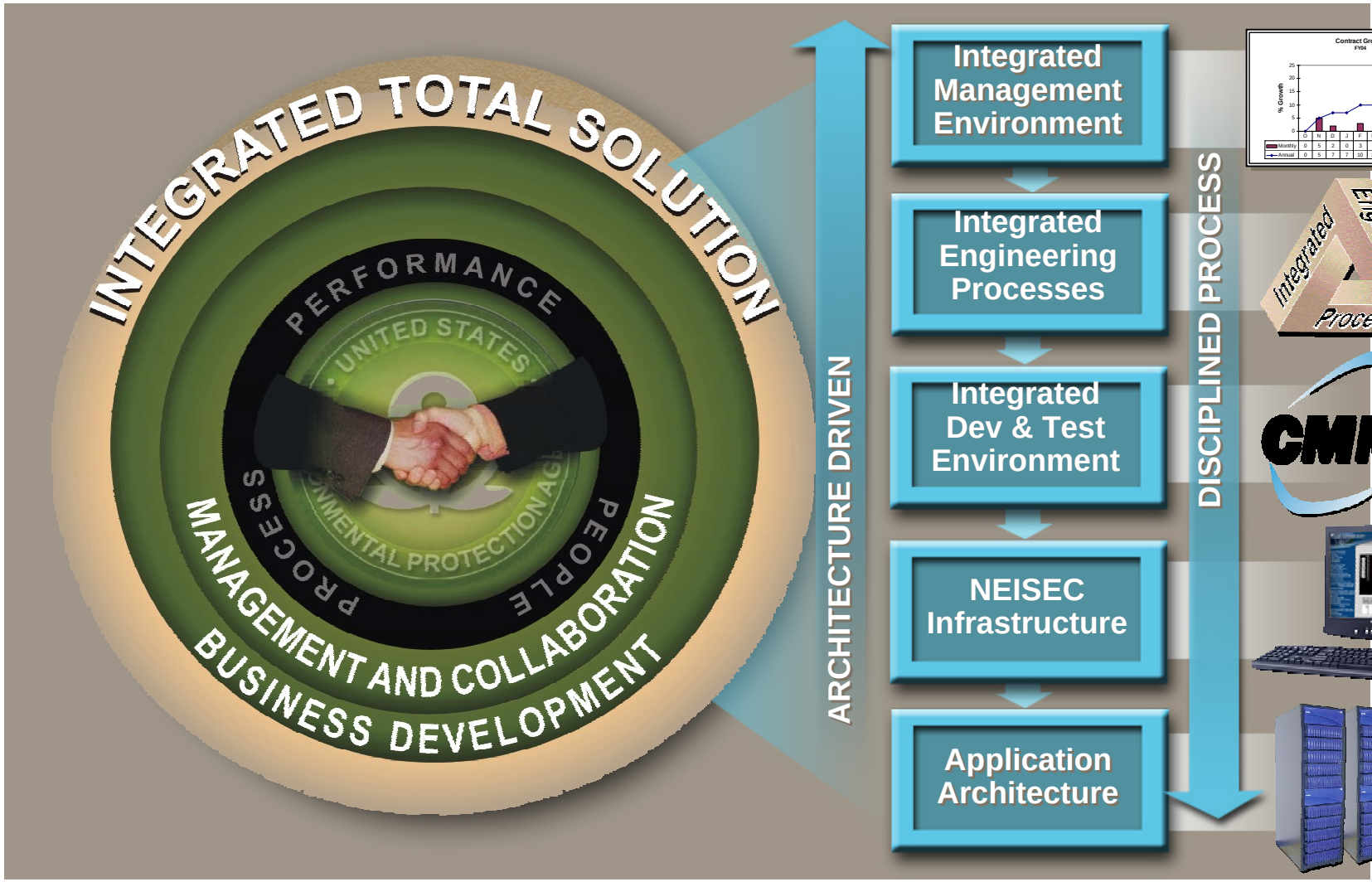
TOPO FORUM

Capability Maturity Model – Integrated (CMMI)

Strategy

Phil Magrogan, CTO
July 2005

Major Components of Lockheed Martin's Integrated Solution

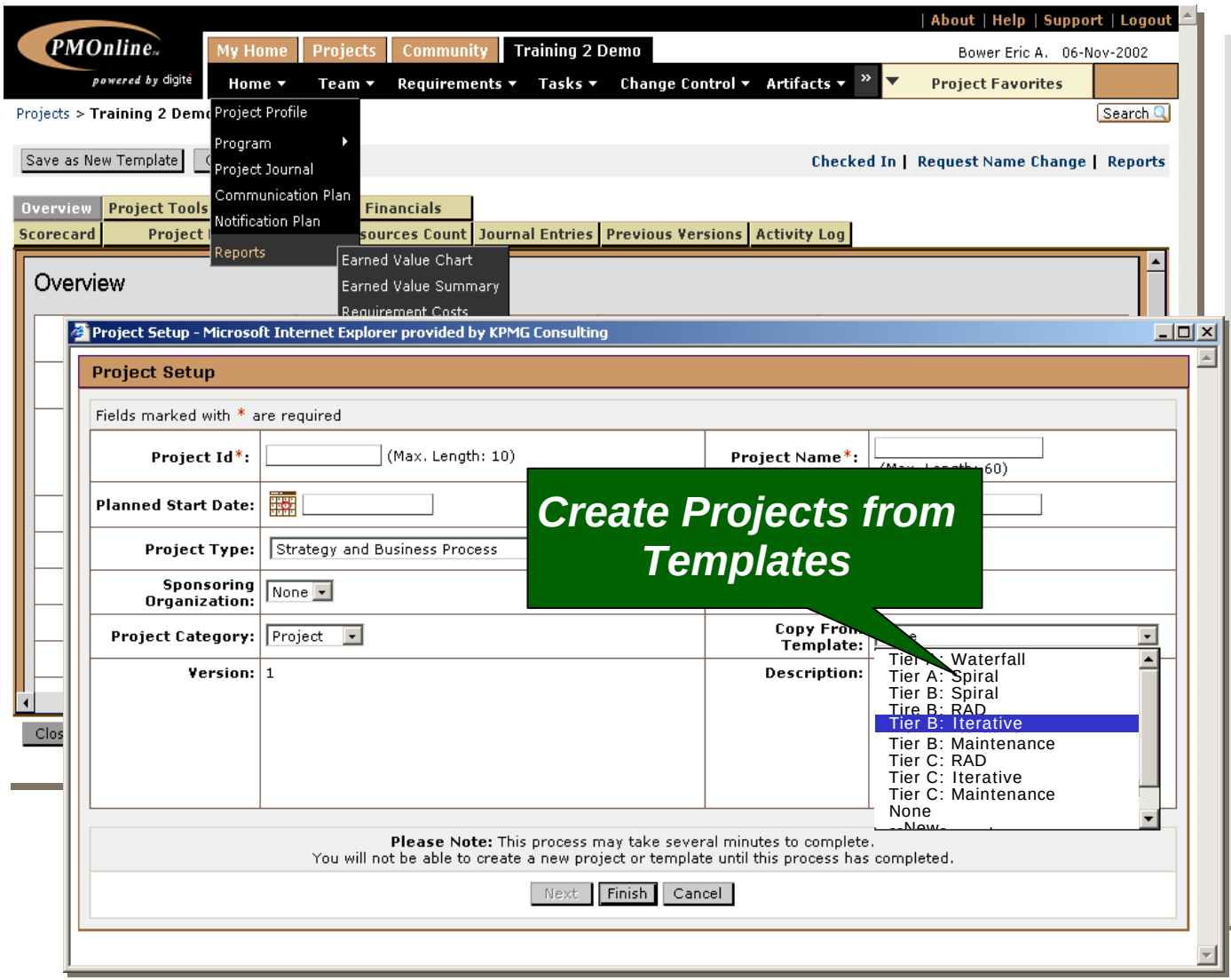


Integrated Project Management Environment

- **Objective: Single access point**
  - Contractual, Task Order, and Project Management Activities
- **ESEconnection.com**
  - Integrates tools for Program and Project Mgt.
    - Controlled access through the NEISEC
    - Contractual Performance
    - Task Order Status & Reporting
  - Public Domain
    - Contract vehicle information
    - Support Ordering Process
    - Communications and Outreach

One stop for current and future Customers

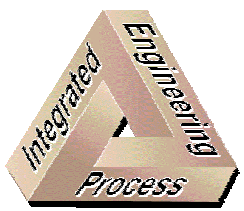
Integrated Project Management Tools

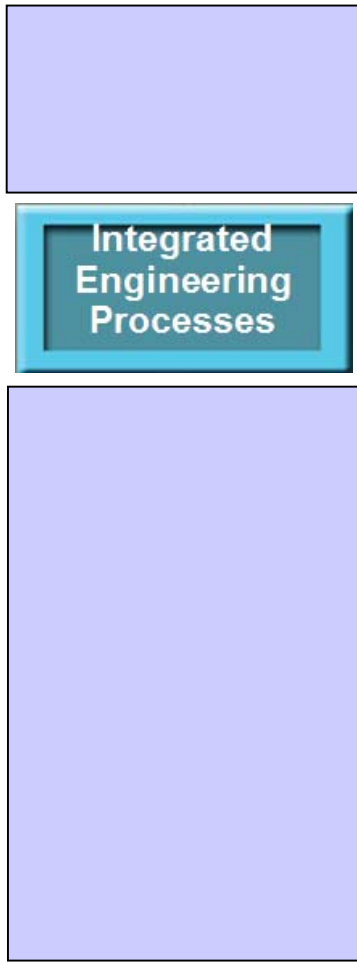


- PMOnline Delivers Collaborative Project Management
- Project Management Tools
- Project Journal
- Communication Plans
- Notification Plans
- Project Reports

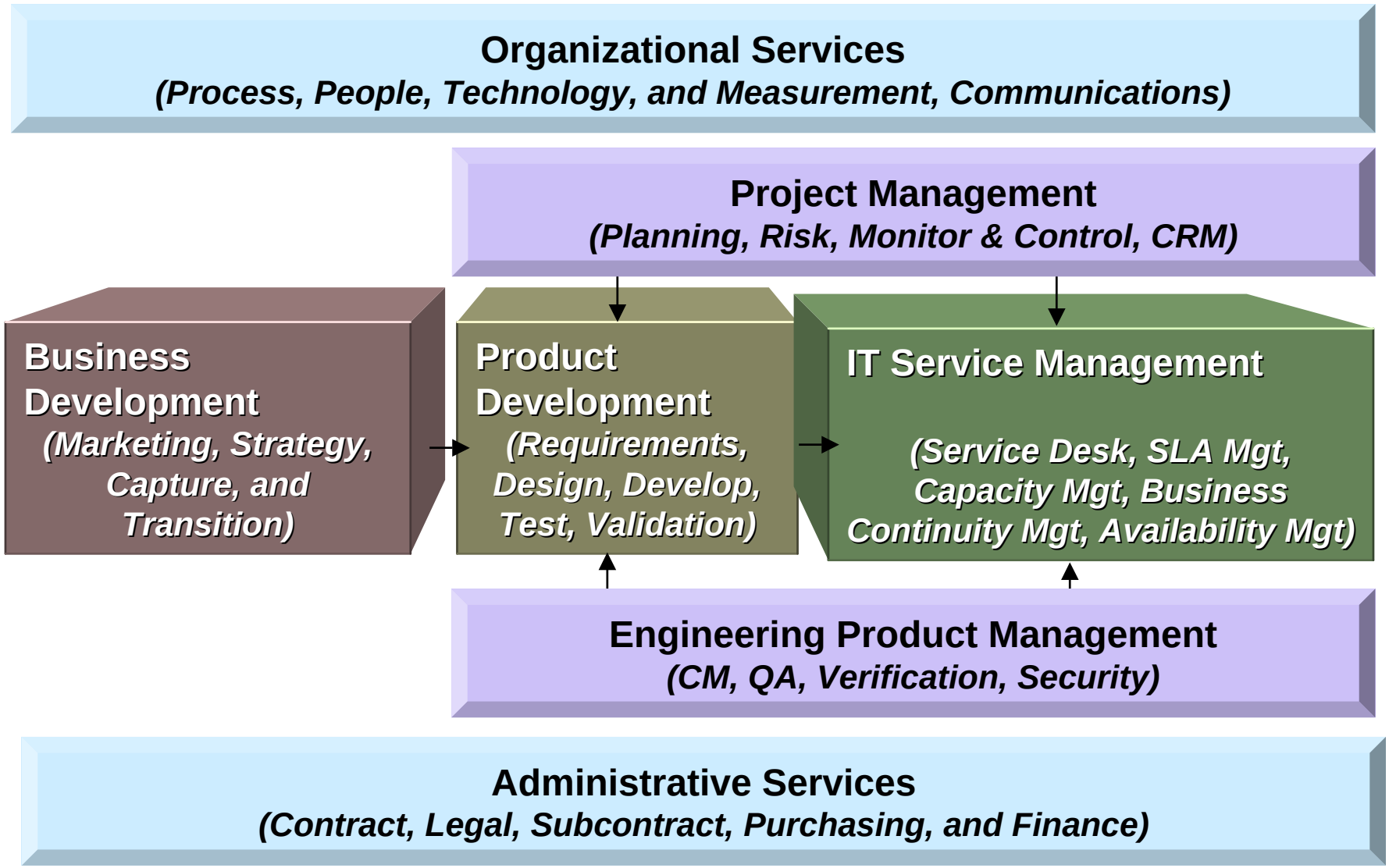
Integrated Engineering Process

- **Apply Sound Systems / Software Engineering practices that:**
  - **Ensure quality product delivery**
  - **Within cost and schedule constraints**
  - **Right-sized application of CMMI compliant processes/practices**
- **Management Oversight & Process rigor determine by work scope**
  - **Effort**
  - **Risk**

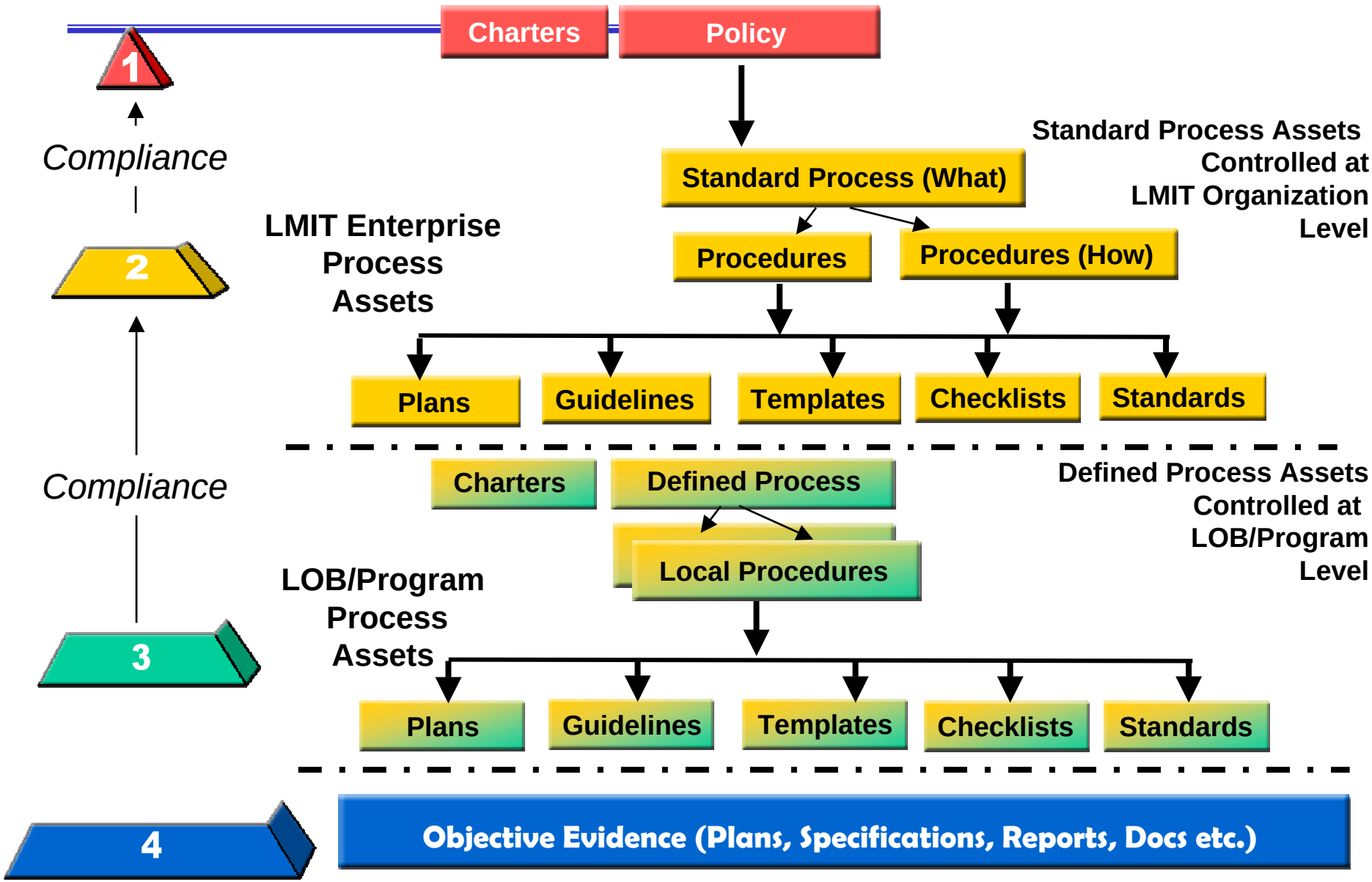




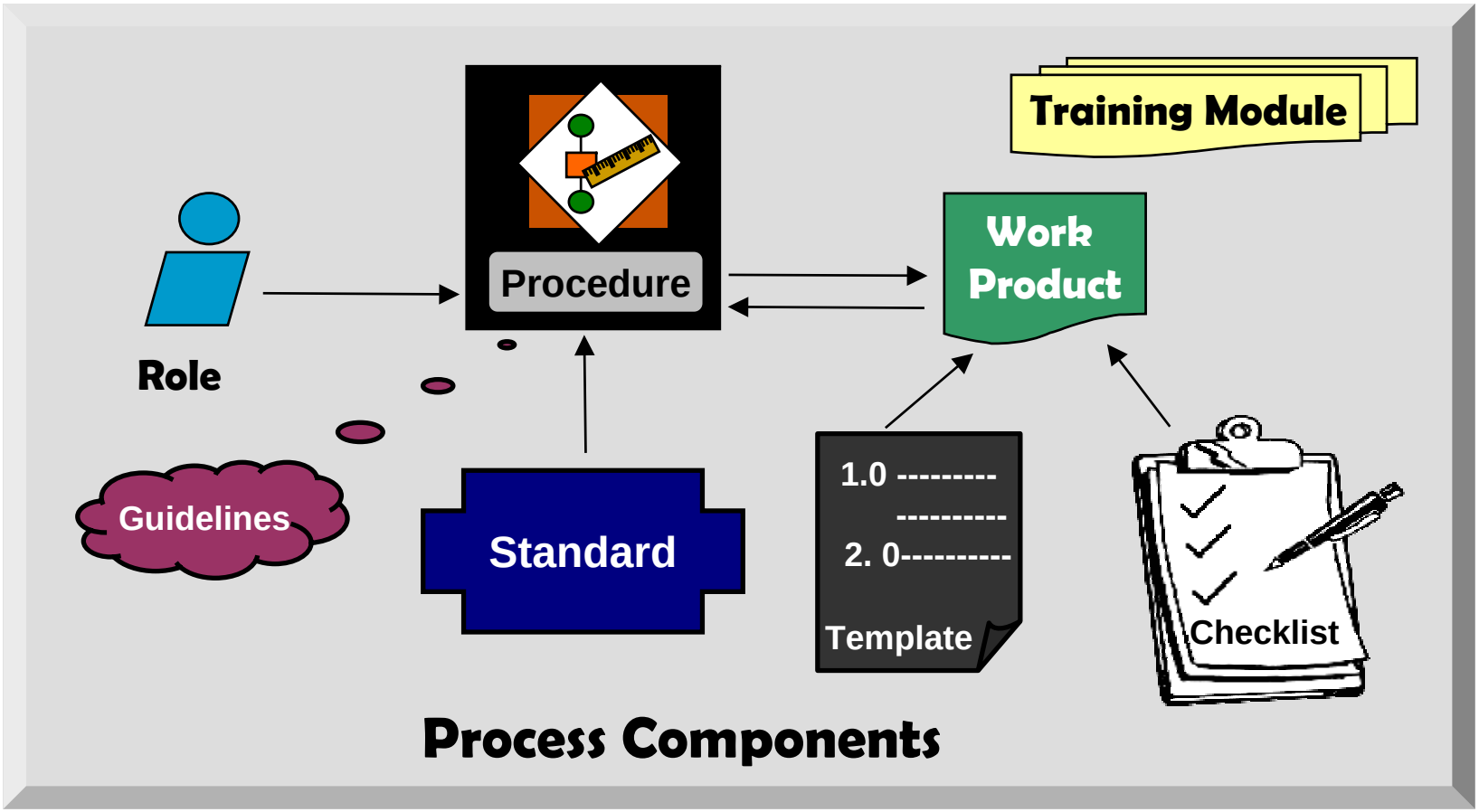
Process Architecture Model



Process Asset Types

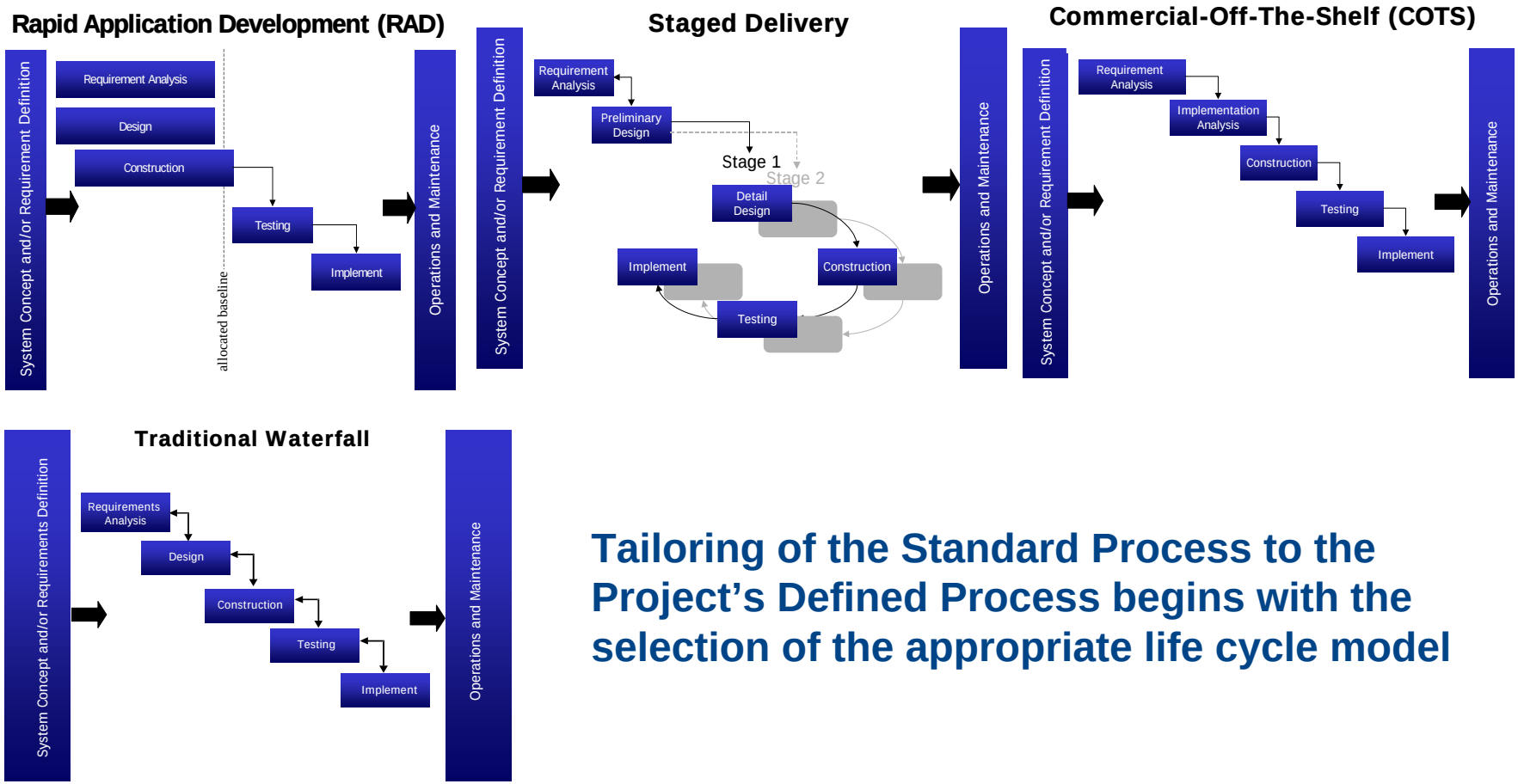


Process Asset Relationships



Life Cycles

Several Life Cycle model alternatives are available

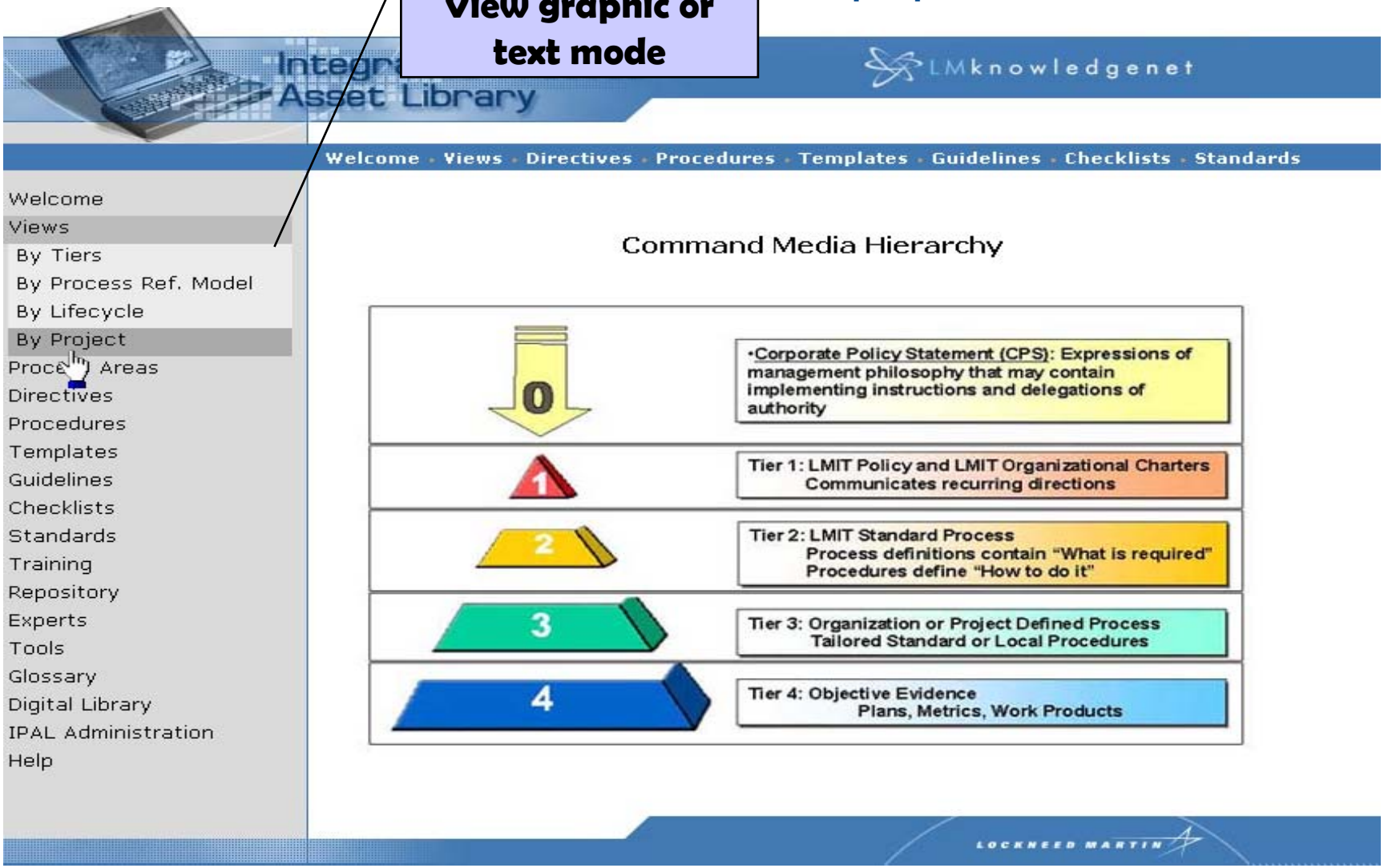


Tailoring of the Standard Process to the Project's Defined Process begins with the selection of the appropriate life cycle model

IPAL Website – WWW Access

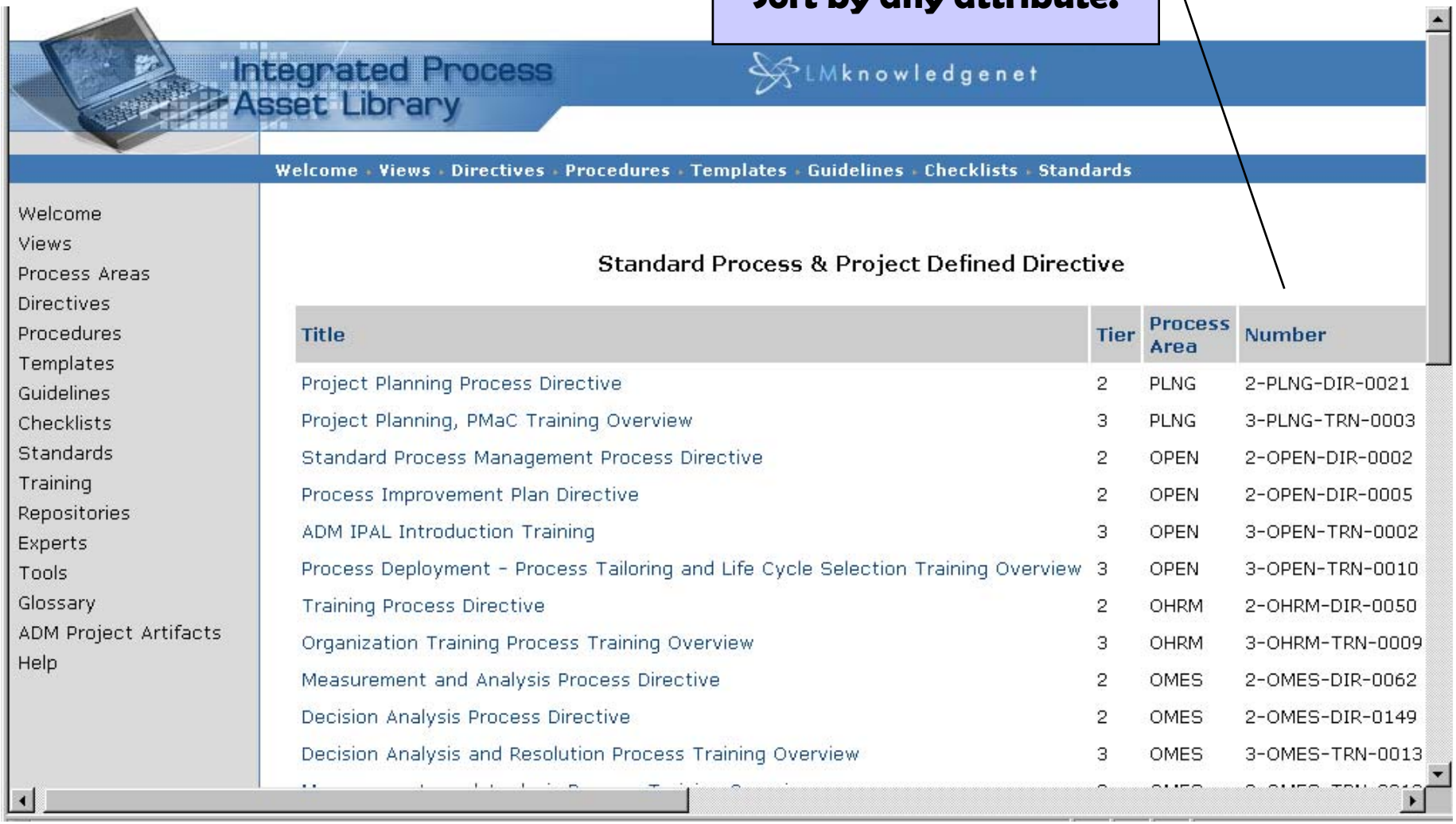
http://ipal.lmit.com

View graphic or text mode



IPAL Site Navigation

Sort by any attribute.



Title	Tier	Process Area	Number
Project Planning Process Directive	2	PLNG	2-PLNG-DIR-0021
Project Planning, PMaC Training Overview	3	PLNG	3-PLNG-TRN-0003
Standard Process Management Process Directive	2	OPEN	2-OPEN-DIR-0002
Process Improvement Plan Directive	2	OPEN	2-OPEN-DIR-0005
ADM IPAL Introduction Training	3	OPEN	3-OPEN-TRN-0002
Process Deployment - Process Tailoring and Life Cycle Selection Training Overview	3	OPEN	3-OPEN-TRN-0010
Training Process Directive	2	OHRM	2-OHRM-DIR-0050
Organization Training Process Training Overview	3	OHRM	3-OHRM-TRN-0009
Measurement and Analysis Process Directive	2	OMES	2-OMES-DIR-0062
Decision Analysis Process Directive	2	OMES	2-OMES-DIR-0149
Decision Analysis and Resolution Process Training Overview	3	OMES	3-OMES-TRN-0013

Tailoring to a Project

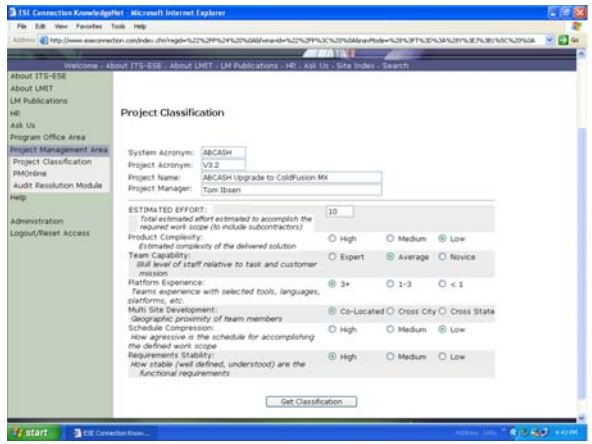
- Select a life cycle model
- Select processes to be included
  - Tailor processes as allowed by tailoring guidelines
- Write Implementing Instruction
  - Extend or replace a Standard Procedure
  - Add contract requirements
  - Add local environment and tool details
- Obtain Organizational Process Group/QA approval

Or, Use pre-defined route maps to save time and improve consistency

Project Classification

Classification Attributes

- **Estimated Labor Months**
  - **Class 1: >60 LMs**
  - **Class 2: >6 and <60 LMs**
  - **Class 3: 0-6 LMs**



- **Weighted values assigned to Scope, Team, & Technology**
  - **Product Complexity** *(High, Moderate, Low)*
  - **Team Capability** *(Experienced, Average, Novice)*
  - **Requirements Stability** *(Mature, Some Instability, High Instability)*
  - **Platform Experience** *(3+yrs, 1-3yrs, <1yr)*
  - **Schedule Compression** *(High, Medium, Low)*
  - **Multi-site Development** *(Co-lo, Cross-City, Cross-State)*

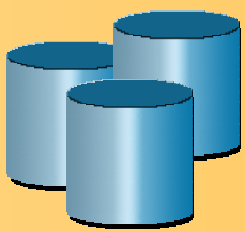
Thresholds adjusted as appropriate

Applying a Framework

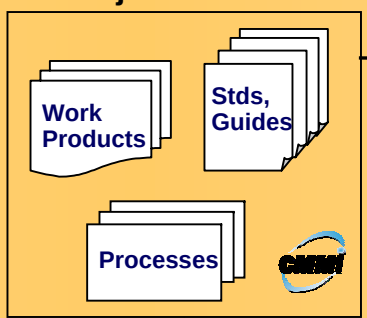
- **Project Classification defined**
- **Project established in PMOnline**
  - **Selects Class Specific Project Framework**
  - **Subset of Process Assets imported**
- **Work Begins**
  - **Detailed planning, SDLC, etc.**

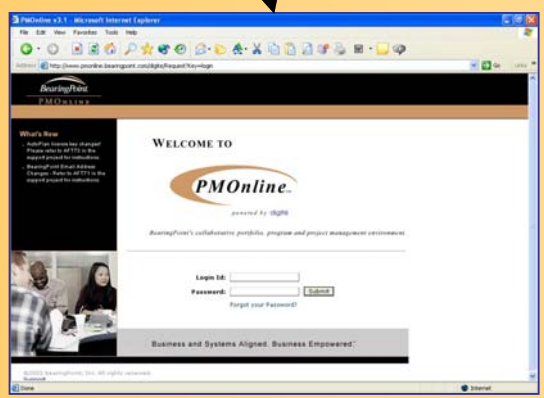
Eliminates Unknowns - focus' team on product development

ESE-PAL Repository



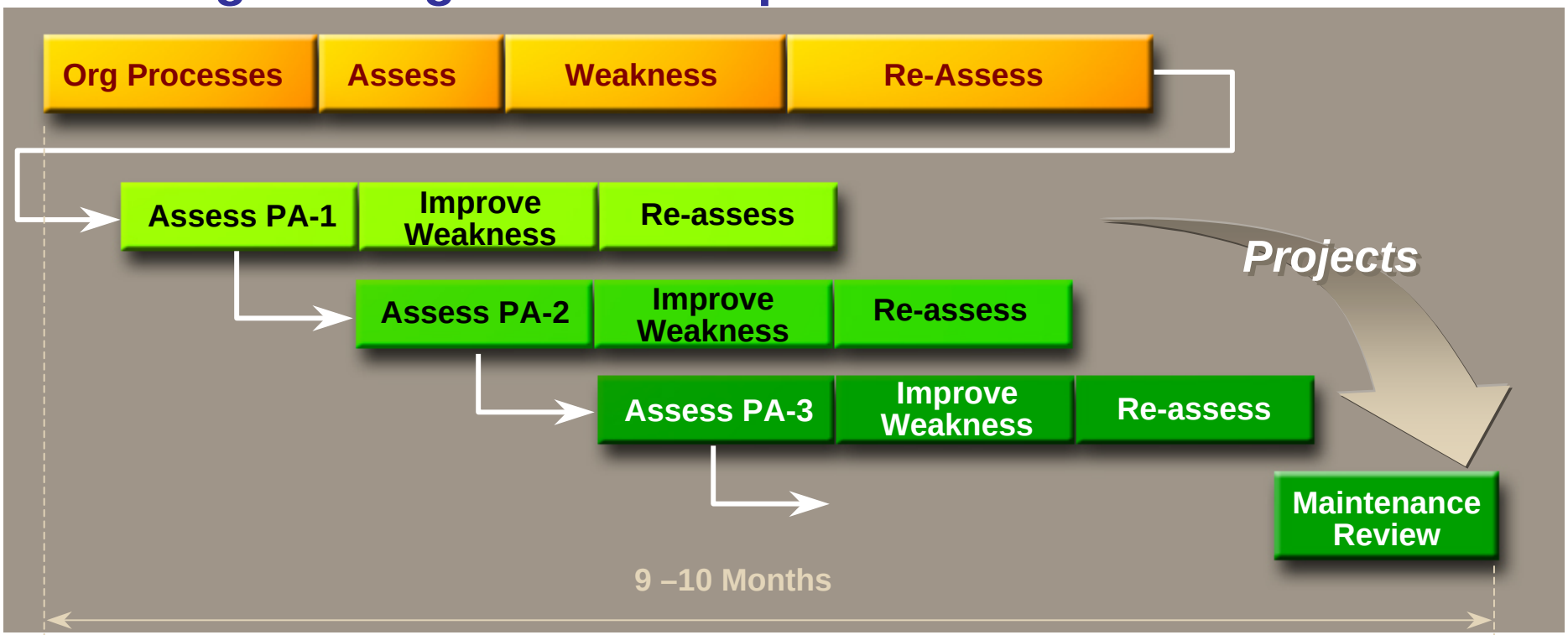
Project Artifacts





Appraisal Innovation - CAM™

- Migrate to CMMI as CMM is Sunset
- Flexible, Non-intrusive, High Quality Results
- Focus on Continuous Systems Engineering Process Improvement



Integrated Engineering Development & Test Environment

- **Common Platform**
  - UNIX for Dev/Deployment
  - PVCS for Configuration Mgt
  - Oracle DBMS and Application Server
- **Technology Focused Toolboxes**
  - Data Mart
    - Informatica, Business Objects, SAS
  - Portal
    - Oracle 10gAS, Oracle Internet Directory
  - Java and &Web
    - Open Source, Eclipse, Dreamweaver, …
  - GIS
    - ESRI
- **Integrated Deployment Automation**
  - Ant, 10gAS, PVCS, . . .

Integrated Engineering Development & Test Tools

Development Environment

- Designer/Forms/Report
- Cold Fusion
- Advantage Gen
- J-Developer
- Verity
- PVCS for CM
- First Logic Address Clean
- Informatica ETC
- Business Objects BI
- XML Schema Builder
- ESRI GIS Tools

Modeling

RAD

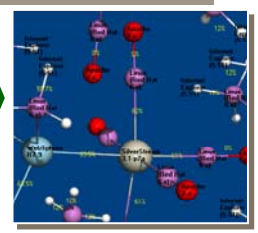
Data Mgmt

Geospatial Integrated Business Intel

Any Source Data

Automated Deployment to all Platforms

Wireless Mobile



Web Services

508 Compliant Portal

APPLICATIONS

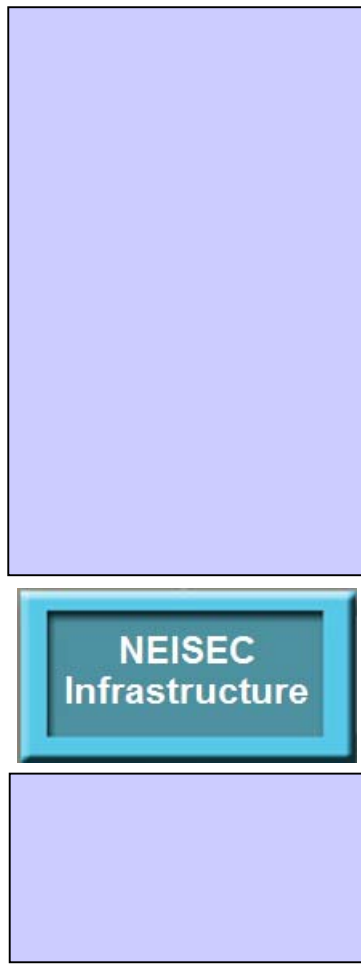
SHARED SERVICES

J2 EE PLATFORM

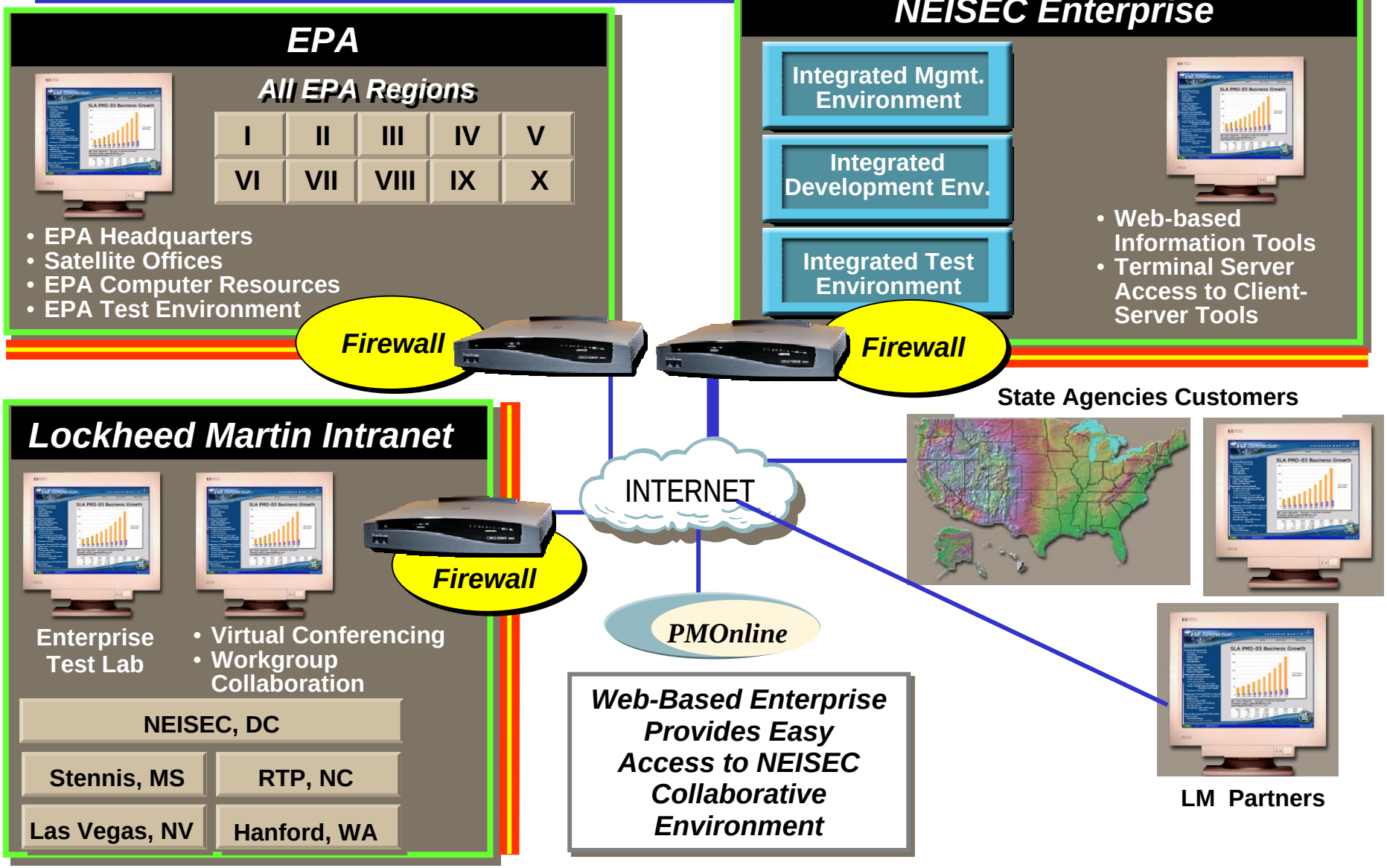
Application Integration

NEISEC Infrastructure

- One System, Multiple Sites
  - Common Databases
  - Shared Tool Licenses
- Web Enabled Access from anywhere
  - Web Tools
  - VPN Access

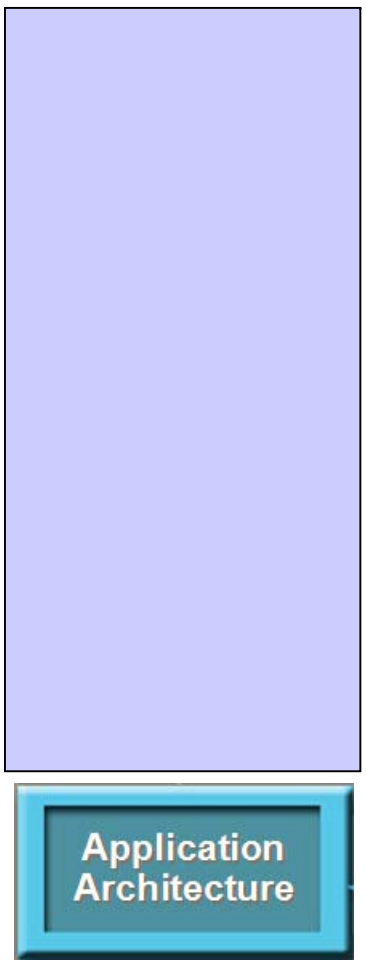


NEISEC Enterprise Environment

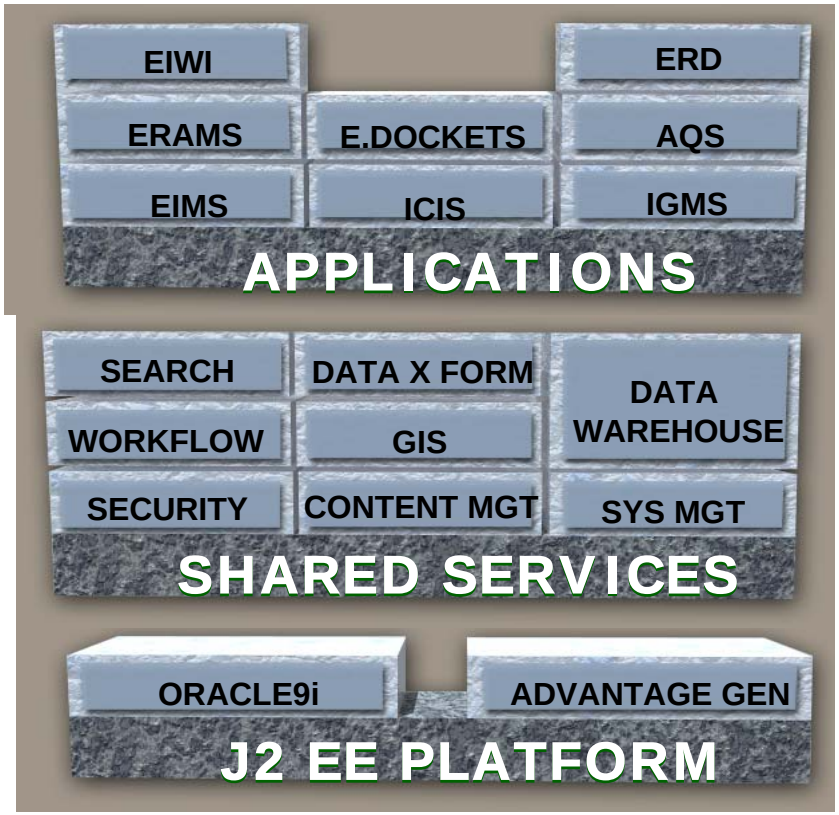


Application Architecture

- **EPA Enterprise Architecture**
  - Service Oriented Architecture
  - Industry Best Practices
  - Open Source Tools and Components
- **Major Components**
  - Portal for Presentation Delivery
  - CDX for secure Exchange Network Partners
  - Common Security and Sign on
  - Common ETL and BI for Metadata management

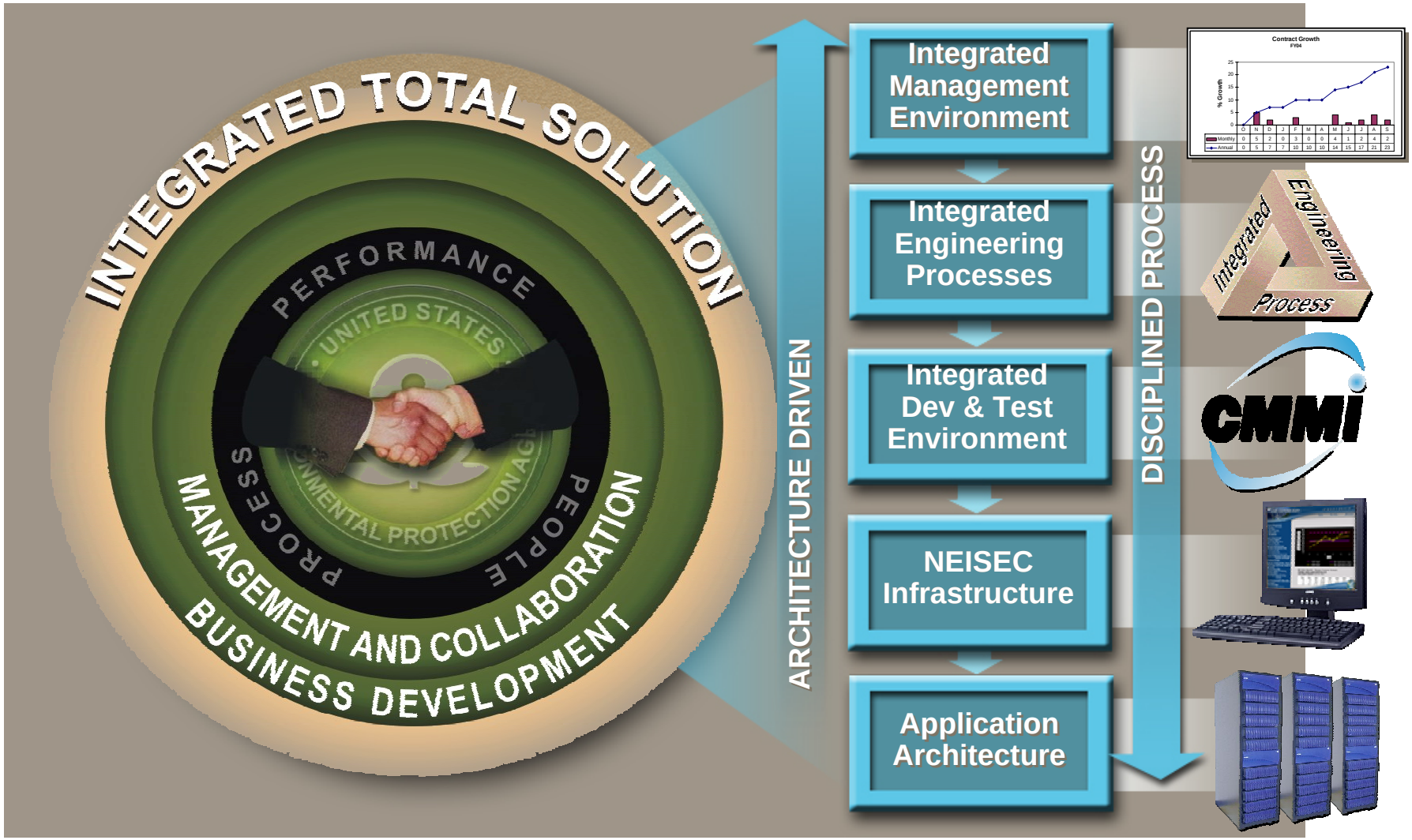


Application Architecture
Service Oriented Solution



- Facilitating Synergy Among Projects
- Collaborating within NEISEC and Partners with Integrated Project Teams
- Utilizing Integrated Development Environment
- Actively Participating on Every Major Standards Body

Summary: Totally Integrated Environment to Achieve CMMI High Matuity, Cost Effective Solution



Facility Registry System (FRS)

…the cornerstone to data integration at EPA

June, 2005

OFFICE OF ENVIRONMENTAL INFORMATION

What is FRS?

The FRS will:

- Create a single, integrated set of facility identification information for each facility
- Provide links to environmental data in EPA's media-specific information systems
- Conform to the *Facility Identification Data Standard* and other applicable standards
- Verify the quality of data, especially the linkages to environmental program information in EPA and State systems

Why is FRS Needed?

- FRS links facility data across program systems, which satisfies EPA's business need to identify:
  - the regulated facilities within a particular sector
  - the location of a facility
  - the different permits and regulations that apply to a particular facility
  - all facilities owned by a particular corporation
- FRS provides essential support for applications that rely on integrated views of facilities:
  - Public Access (Envirofacts, EnviroMapper)
  - Enforcement (IDEA, OTIS)

FRS Features

- Single source of comprehensive and authoritative information on regulated facilities
- Centrally managed database identifying EPA-regulated facilities
- Agency-wide data integrator using the facility identifier
- Identification and correction of data gaps, errors, and duplication of facility information
- Comments and corrections sent directly to EPA; improves the reliability of the facility information (Integrated Error Correction Process)

FRS Features

(continued)

- State Facility Identification Management System database (South Carolina, Arizona, North Carolina,California, Indiana, Iowa, Missouri, others)
- Geocoding capability for facility locational data
- Precise verification and management procedures to reconcile and integrate information
- Comprehensive search capabilities

Facility Identification Background

- February 1998 -- EPA issues *Facility Identification Interim Data Standard*
- October 1998 - Facility Linkage Application replaces FINDS
- October 1998 - Reinventing Environmental Information (REI) establishes a revised set of goals for facility identification
- July 1999 - EPA issues *Proposed Facility Identification Data Standard Final Draft*, developed by State/EPA Action Team
- September 1999 - Facility Registry System (FRS) Beta Version completed
- December 1999 - New team from OIC is formed

Facility Identification Goals
(Approved by REI Subcommittee 10/7/98)

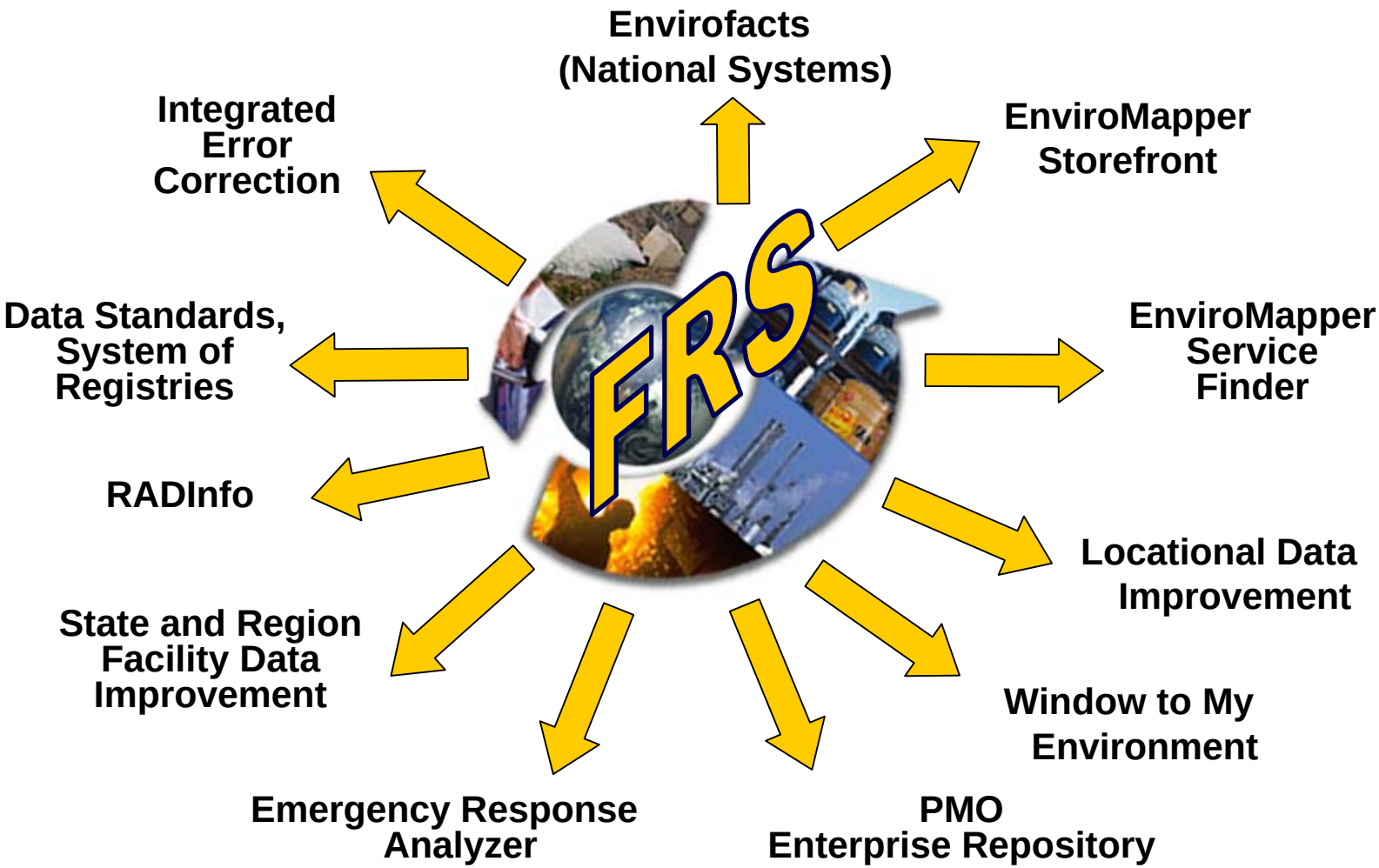
- Create an authoritative, up-to-date source of facility data for EPA, the States, and the public.
- Improve the quality of facility identification data.
- Provide an “anchor” to link facilities, manage electronic reporting, and exchange integrated data with States and industry.
- Reduce the reporting burden on the regulated community.

FRS Successes

- 1.6 million unique facility records linking to over 2.0 million Program IDs
- First project to exchange data with States through CDX for incoming and outgoing data
- Integration of 28 State master files
- Error Correction
  - Over 13,000 error notifications over 3 years
  - 12,750 errors resolved to-date
- Data quality improvement
- Designation and management of over 7,800 federal, 50,000 tribal land, and over 300 National Environmental Performance Track (NEPT) facilities

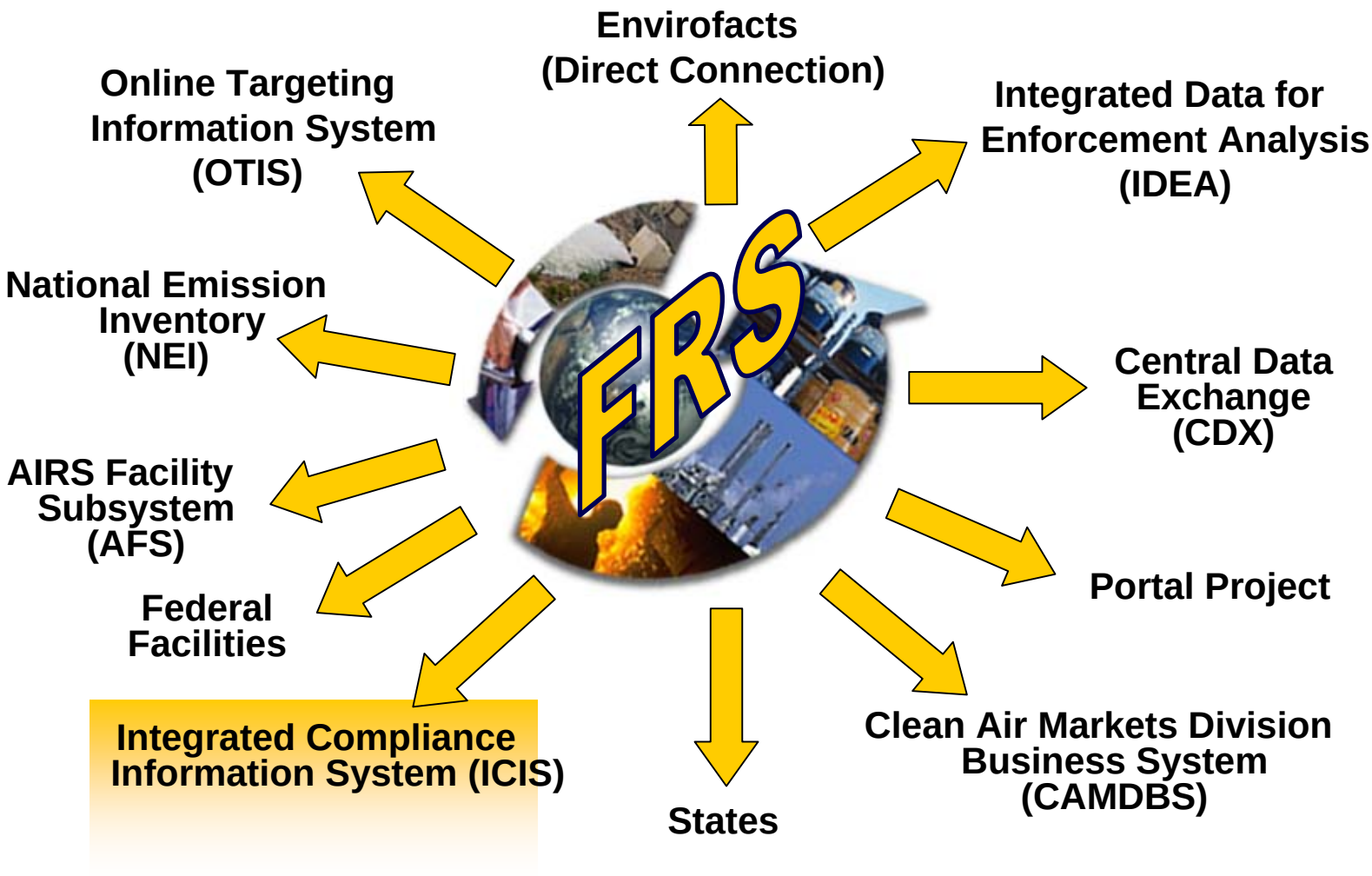
FRS Dependencies

FRS is critical to the successful operation of these activities!



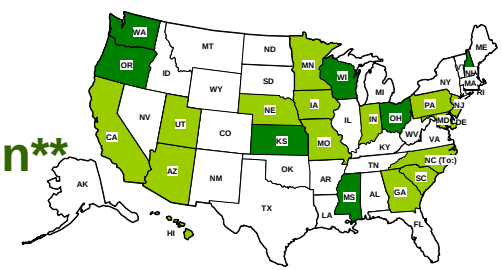
FRS Dependencies

These activities rely on FRS data integration!



Key FRS Activities

- National Systems reconciled on a regular basis*:
  - TRI*
  - RCRAInfo*
  - RMP*
  - TBA
  - PCS-Majors & Minors*
  - NEI-1999
  - NEPT*
  - CAMDBS
  - CERCLIS*
  - ICIS*
  - RADInfo
  - NCDB*
  - BMS
  - AIRS/AQS
  - AIRS/AFS*
  - SDWIS*
- State exchanges from:
  - Pennsylvania
  - South Carolina
  - **Mississippi****
  - Utah
  - Minnesota
  - Arizona
  - **Washington****
  - Indiana
  - **Ohio****
  - Nebraska
  - Maryland
  - Iowa (To:)
  - New Jersey
  - North Carolina (To:)
  - California (Solid Waste Program)
  - Missouri (To:)
  - **New Hampshire****
  - Hawaii
  - **Oregon****
  - Georgia
  - **Wisconsin****
  - **Kansas****

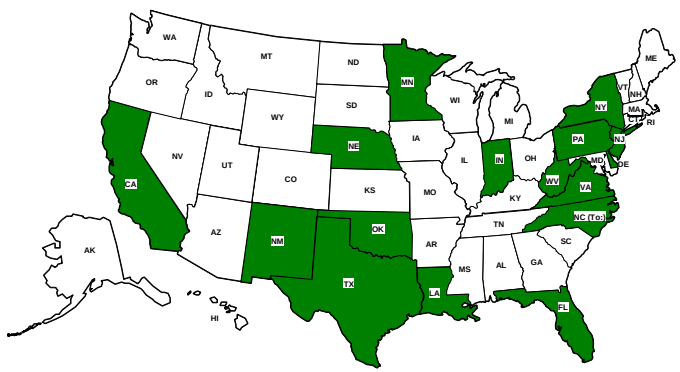


**** - Exchange is through the Network Node.**

OFFICE OF ENVIRONMENTAL INFORMATION

Progress

- State exchanges through Network Node from:
  - New York
  - New Jersey
  - New Mexico
  - Delaware
  - Texas
  - Florida
  - West Virginia
  - Louisiana
  - Pennsylvania
  - North Carolina
  - Nebraska
  - Indiana
  - Minnesota
  - California
  - Virginia
  - Oklahoma



OFFICE OF ENVIRONMENTAL INFORMATION

Challenges

- Conformance to data standards
- Consistency in the national database population
- Data quality
- Data gaps
- Feedback to system owners

Self-Reported Linkages

- Relying *only* on self-reported linkages raises the possibility of two types of errors:
  - The linkage may be reported incorrectly by the facility
  - The linkages may not have been reported by the facility, when they should have been

Self-Reported Linkages - Results

- 28,275 linkages reported by TRIS facilities
  - 23% not found in Envirofacts
- 5,110 linkages reported by RMP facilities
  - 49% not found in Envirofacts
  - RMP did not report the information system associated with the linkage

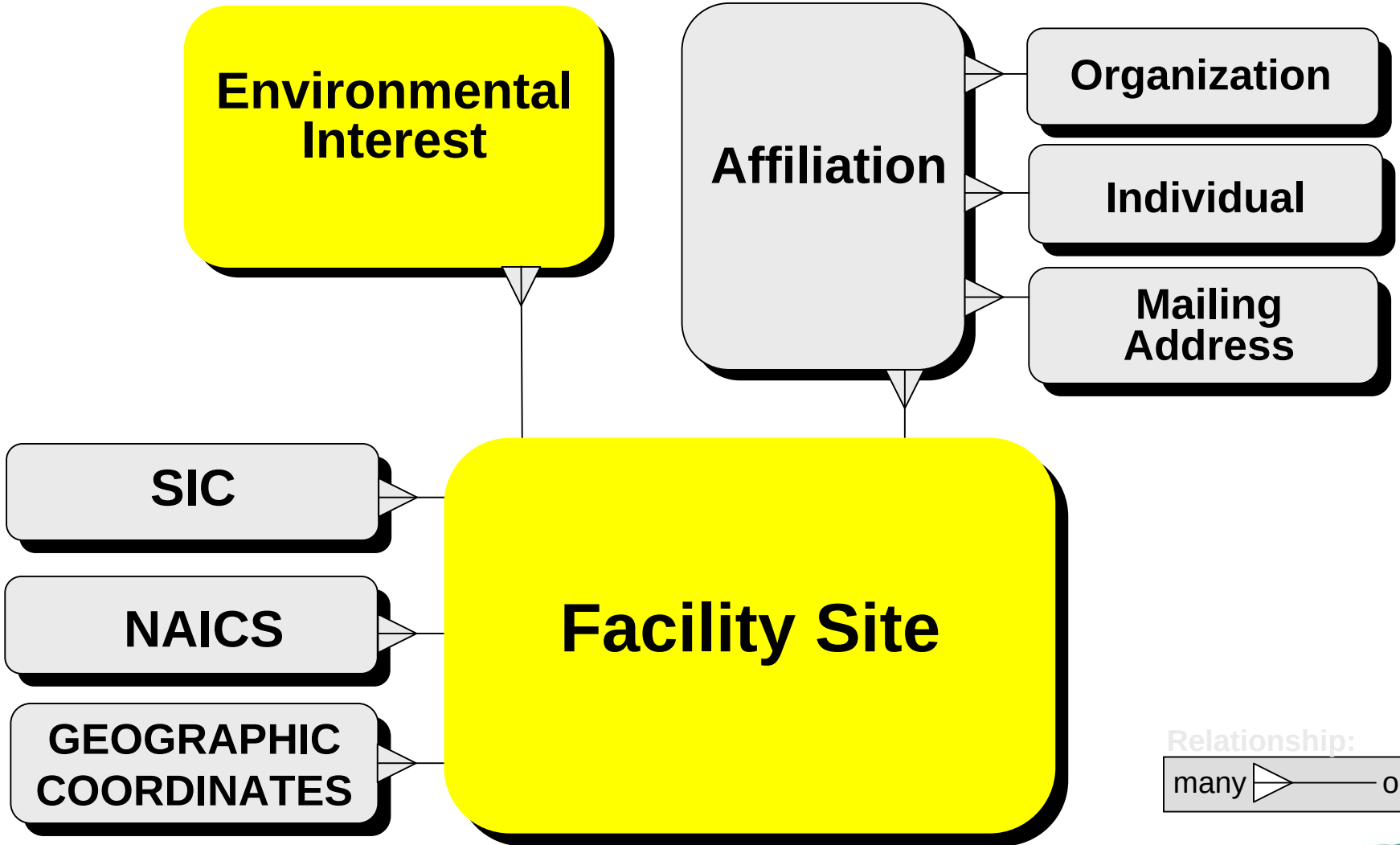
Technical Session

FRS Requirements

- Data requirements
- Functional requirements
- Security requirements

Data Requirements

FRS High Level Data Model



FRS Data Sources

- Meeting with programs to verify the mapping of program facility data to FRS
- Available on FRS Web site for review (under development)
- Initial data sets
  - TRI 1998 & 1997 reporters - 25,539 facility sites
  - RMP (as of February 2000) - 14,753 facility sites

FRS Initial Data Sets

- Initial data sets - continued
  - RCRIS TSDs - 4504 facility sites
  - PCS Majors - 6,893 facility sites
  - BRS 1997 & 1995 reporters - 20,982 facility sites
  - State Master facilities - To be selected
- Data will be updated on a regular basis

Facility Site

- All Data relationships currently tracked by FLA will be maintained
- A new FRS Facility Site record will be created
  - Data may be obtained from multiple sources
  - Data source and latest update date will be tracked by data element
  - Some data will be derived
  - Data will be standardized and will conform to facility identification standard

Environmental Interest

- Maintain linkage numbers to EPA program and State environmental information systems
- Environmental Interest Types, examples:
  - Hazardous Waste Handler - TSD
  - Hazardous Waste Handler - LQG
  - NPDES Major
- Alternative and Historical Names
- Alternative Identification (Support FITS)

Environmental Interest Start & End Dates

- Environmental Interest Start Date
  - When did the program first become interested in this facility site?
- Environmental Interest End Date
  - When did the program cease to have an interest in this facility site?
- Start Date Qualifier, example:
  - ‘FIRST REPORTING YEAR’
- End Date Qualifier, example:
  - ‘DATE LAST SUBMISSION RECEIVED’

Affiliation

- Affiliation Start & End Dates
- Affiliation Types
  - Regulatory Contacts - maintain most current contacts per program
  - Owner/Operator - applicable to program only or applicable to enterprise level?

Functional Requirements

Create FRS Record

- Extract data
- Standardize data for matching
- Parse address
- Evaluate data quality
- Process data
- Complete FRS record

Standardize Data for Matching

- Standardize facility name
  - Remove special characters (- -,_, ., @, /, (, +)
  - Replace individual words with standardize words (e.g., Assoc. → Association, Chem → Chemical, Elec → Electrical)
  - Drop certain components (e.g., Inc, Company, Corporation)

Standardize Data for Matching *(Continued)*

- Standardize location address
  - Remove special characters
    (- -,_, ., @, /, (, +)
  - Replace individual words with standardized words
    (e.g., Ave → Avenue, Rte → Route, Hwy → Highway)
  - Replace state names with state abbreviations
  - Replace spelled out numerals to their abbreviated form
    (convert Twenty First to 21st)

Standardize Data for Matching *(Continued)*

- Standardize City and County names
  - Remove special characters from the name (- -,_, ., @, /, (, +)
  - Replace individual words with standardized words (e.g., Ctr → center, Spg → Spring)
  - Remove buzz words (e.g., ‘City of’, ‘Town of’, ‘South of’)
  - Distinguish between CO as County (Fairfax Co) and CO as Colorado (Co Springs)
  - Distinguish between LA (Los Angeles) and LA PLATA (La as an article in Spanish named cities)

Parse Address

- Parse standardized location address into key words, street number and street name
  - Isolate key words and the number following the key word
    Examples: HWY , PR, INTERSTATE , UNIT, BLK, LOT, ROUTE, UNITED STATES, ROAD,SR, POBOX, CR, CNTY ROAD, BLDG
  - Parse the standardized location address minus the key words to obtain a street number and a street name

Data Example

- Raw data for Program –1

  Name = Congressional Schl

  Location Address = 6565 Arlington Blvd, (Route 50 W)

  City = City of Falls Church

  County = Fairfax County

  State = VA

  Zip Code = 22042

- Processed data for Program –1

  Standardized Name = Congressional School

  Standardized Location Address = 6565 Arlington Blvd Route 50 W

  Parsed_Street_number = 6565

  Parsed _Street name = Arlington

  Street suffix = Blvd

  Street pre-directional code = W

  Key word = Route 50

  Standardized city name = Falls Church

  Standardized county name = Fairfax

Evaluate Data Quality

- Core data elements are evaluated and assigned a data quality code

Facility Site Name	Location Address
Locality Name	County Name
State Code	ZIP Code

- Data quality code identifies
  - Missing data values
  - Invalid data values
  - Address Type

Evaluate Data Quality *(Continued)*

- Standard City Name, County Name, State Code, and ZIP Code checked for consistency using:
  - FIPS 55-DC3
  - USPS ZIP Code Reference Tables
- If the combination of Standard City/Standard County/State/Postal code is invalid, the invalid element(s) are identified
- Certain name anomalies are identified as invalid (e.g., ‘UNKNOWN’, ‘NOT GIVEN’, ‘NO ADDRESS GIVEN’)

Evaluate Data Quality *(Continued)*

- Address Types
  - Urban – contains a street number and a street name
  - Directional – contains directions
  - Mailing – contains P.O. Box number
  - Irregular – missing street number or street name or both and is not in any of the 3 types listed above

Process Data

- Automated matching process
  - Valid data quality code
  - Data quality code with an invalid or missing data element which can be derived
- Flagged for ‘Manual Review” (unless already linked by Data Stewards)
  - Invalid or missing data that cannot be derived
  - Directional, Mailing, or Irregular Address Types
  - Use available sources to fill the data gaps

Automated Matching Process

- Scoring processes
  - **City Name , County Name, State and Zip Code MUST match Score = 50.**
  - ***IF*** **Parsed Street Number and Parsed Street Name match Score = 25.**

    OR **Parsed Street Number and Key_word* (or Key_word equivalency) match Score = 25**
  - ***IF*** **Standardized Name match Score = 25**

    OR **Alternate Name match Score = 25**

* Key_word equivalencies are used for match-up
(example: Route, SR, CR → Highway)

Automated Matching Process

(Continued)

- Evaluation Processes
  - If Score = 100, only Environmental Interest record is established
  - If Score is less than 100
    - Make use of linkages established by Data Stewards
      - Create new FRS record
      - Establish Environmental Interest
      - Flag for Manual Review

Manual Review

- Fill in data gaps on FRS record
- Create manual linkages
- Review stand alone facilities
- FLA Statistics

FACILITIES FLAGGED FOR MANUAL LINKAGES IN FLA By Program/Region

Region	PCS Count	RCRIS Count	RMP Count	TRIS Count
01				
02	34	103		1
03	60	28	19	55
04	22	9		1
05	79	34		2
06				
07	289	183	98	13
08	87	32	4	3
09				
10	127	255	44	12

Search

- Name (AKA)
- Location Address
- Environmental Interest/Type
- SIC/NAICS
- Organization

Reports

- Customized Reports
- Facility Reports

Linkage Verification and Lock Process

- FRS will verify and lock linkages
- Locked linkages can only be unlocked through
  - Error Correction Process
  - Discrepancy Reporting
  - Primary Data Steward

Security Requirements

- APPLICATION SECURITY REQUIREMENTS:
  - User verification and certification
  - Client/Server security

Major Issues

Basic Assumption:
Central management of facility data.

- Data Flow
- Acceptance of State Data
- Reconciliation Process

FRS Next Steps

- Continue scheduled updates to support applications that rely on integrated views of data (Envirofacts, EnviroMapper, IDEA, OTIS)
- Continue to support Data Steward clean-up efforts
- State Pilot efforts that support central receiving

FRS Next Steps - Short Term

- Resolve duplicates:
  - Resolve remaining cases on a case-by-case basis through coordination among Program Offices
  - Fix cases where there has been an error in the TRIS or RMP database assignment of a facility identifier, possibly stemming from dissimilar name and address data

FRS Next Steps - Long Term

- Data Flow:
  - Existing program system extracts
  - Central receiving registration
  - State master records
- Acceptance of State data:
  - Model MOU for data quality

FRS Next Steps - Long Term

- Reconciliation Process
  - Establish automated and manual verification and error correction processes involving the facility and/or States
- Populate FRS with additional facilities from:
  - RCRA, TSDs and LQGs
  - PCS Majors
  - AIRS/AFS Majors
  - State master records
- Build central receiving and electronic reporting capability

Contracting for IT Services

Sanjib Chaki, Steve Hufford, Dwayne Aydlett

OFFICE OF ENVIRONMENTAL INFORMATION

EPA's IT Services Contracting Strategy

- Create contracts specific to lifecycle stage:
  - Conceptual design, requirements development, architectural planning (ITS-BISS)
  - Detailed design and development, operations and maintenance, and enhancement (ITS-ESE)
  - Production operations, networking, and IT infrastructure (ITS-EPA)
- Perform contract administration/oversight centrally, and offer service throughout EPA
- Ensure sound technical performance and cost-effective services

ITS-BISS: Orientation

- Performance-Based Contract that emphasizes linking ITM to management and environmental results
- Multiple Award Contract that requires multiple contractors to compete for work
- Customers have choice of various Task Order Types: Fixed Price, Cost-Plus, Time & Materials
- Allows to choose the best among good, innovative solutions developed by the contractors.
- Competes Task Orders on price, performance factors & technical approach.
- Provides best value & better performance.
- Provides tools to help achieve project success:
  - Performance management
  - Quality assurance
  - Incentives to encourage excellent contractor performance

ITS-BISS: Contract Objectives

BISS contractors also are expected to perform projects in a manner that…

1. Advances Agency's five mission goals.
2. Strengthens Agency's capacity to meet management goals.
3. Leads to better business and ITM alignment, information leadership, analytical capabilities, governance, and information technology services.
4. Grows the value and usage of the contract to customers, stakeholders, and partners through effective solutions partnering.
5. Establishes a rapid response capability that delivers solution-oriented advice and assistance on critical ITM policy, programmatic, and marketplace issues.
6. Leverages Agency's ITM policies, including EA, Investment Management Process, Security, Data Standards, etc., to improve mission and management outcomes.
7. Implements contract management practices that facilitate partnership and innovation between contractor and the Agency.

ITS-BISS: Contract Details

- Awarded in April 2005
  - Two year base period, with two options
  - Total 5 years
- One Full and Open (Large) Contractor and One Small Disadvantage Business
- Contract Capacity for Large Contractor is ~$150M
- Contract Capacity for Small Disadvantage Contractor ~$125M
- Awarded 14 Task Orders of various types in less than two months.

ITS-BISS: Major Functional Areas

- ITM Policy
- ITM Planning
- ITM Security
- ITM Investment Management Advice, Assistance & Support
- Enterprise Architecture
- Intra-and-Inter-Governmental ITM Initiatives Support
- ITM Acquisition Support
- Conference Support & Meeting Facilitation
- Independent Validation and Verification

ITS-ESE: Orientation

- $700M ceiling, 9 year potential lifespan
- Wide variety of professional application systems engineering services
  - Detailed design and development, operations and maintenance, and enhancement
  - Data and metadata management
  - Applications security support
  - Training, GIS, statistical analyses, etc.
- Nationwide support to USEPA and partners
- Performance-based contract
- Managed primarily at the Task Order level

ITS-ESE: Contract Objectives

- Develop a full partnership relationship with the Offeror
- Apply CMM-SW or CMM-I to reduce development costs, expedite completion of small, low risk projects, and minimize risk in large, complex projects.
- Become the Agency's and partners' vehicle of choice
- Implement industry best practices
- Obtain an integrated solution for EPA, with strengths of the Offeror and specialized 'best of breed' companies.
- Maximize project synergy regarding data and technology standards, enterprise architecture, and systems interfaces.
- Help EPA meet its socio-economic procurement goals.
- Maintain the ability to support audits by EPA and Federal oversight groups

ITS-ESE: Contract Details

- A protest of the contract award significantly hindered smooth transition from prior contract
- Performance-based contracting is difficult to implement for software engineering
- Connecting the System Engineering Center securely to EPA's computing infrastructure is more challenging than anticipated
- Cross-project synergy is possible, although it requires focus and attention
- Customers do not always think first of following enterprise directions

ITS-EPA: Orientation

- Task Order on GSA FEDSIM Millennia contract
  - Estimated value $867M over 7 Years
  - Period of performance (annual periods) 09/04/02 : 09/30/09
  - Hybrid Contract Type
  - Performance Based, moving towards Managed Services

ITS-EPA: Contract Objectives

- Key objectives of ITS-EPA
  - Partnership resulting in cost reduction
  - Faster time to implement new technology
  - Shift of OTOP employees focus to customer service
  - Utilize the capabilities of specialized best of breed firms
  - Leverage vendors best practices

ITS-EPA: Contract Details

- Performance-Based
  - Requirements described in terms of results required
  - EPA orders service not labor hours
- Project Management Plans
  - Primary means of communicating the “what, how, when, & where” of the CLIN

OFFICE OF ENVIRONMENTAL INFORMATION

ITS-EPA: Tasks/Functions

- Call Center
- Campus Network Management
- Desktop Support
- IT Training Institute
- Voice
- High Performance Computing (HPC) Operations
- Email and Lotus Applications
- Wide Area Network (WAN)
- Web and Application Hosting (WAH)
- Facility Operations
- Image
- Enterprise Server
- Geographical Info. Sys. (GIS)
- Distributed Systems
- Technical Consulting
- Local Technology Services
- Internet Service Center
- Emerging Requirements
- General Functions

OFFICE OF ENVIRONMENTAL INFORMATION

ITS-EPA: Managed Services Transformation

- Definition
  - What is the difference between Performance Based Contracting & Managed Services?
    - PBC = Defined Outcomes, Metrics, Incentives, Quality Assurance Plan
    - MS = PBC + Contractor Provides ALL resources
      - Personnel, Equipment, Facilities
  - EPA defines service requirements and performance expectations
- Benefits
  - Contractor performs to meet service levels
  - Potential for cost savings through use of shared resources

Open Discussion

OFFICE OF ENVIRONMENTAL INFORMATION

Toxics Release Inventory

Larry Reisman
TRI Program
reisman.larry@epa.gov
(202) 566-0751

Toxics Release Inventory

Toxics Release Inventory Reporting Requirements (Section 313 of EPCRA and Section 6607 of PPA) – *Do I Need to Report*?

Toxics Release Inventory

WHO MUST REPORT?

- **Facilities in covered primary SIC code(s) and federal facilities; and**
- **With 10 or more full-time employees (equivalent of 20,000 hours per year); and**
- **Chemical-by-Chemical Determination**
  - **That exceed "manufacture", and/or "process", and/or "otherwise use" thresholds for each Section 313 chemical.**

Toxics Release Inventory

- **COVERED SIC CODES**

Industrial Sector	Primary SIC Code
Manufacturing	20-39
Metal mining	10 (except 1011, 1081, and 1094)
Coal mining	12 (except 1241)
Electric utilities	4911, 4931, and 4939, limited to facilities that combust coal and/or oil for the purpose of generating electricity for distribution in commerce
Treatment, Storage, and Disposal facilities	4953, (limited to facilities regulated under the Resource Conservation and Recovery Act, Subtitle C, 42 U.S.C. Section 6921 *et seq*.)
Solvent recovery services	7389, limited to facilities primarily engaged in solvent recovery services on a contract or fee basis
Chemical distributors	5169
Petroleum bulk terminals	5171

Toxics Release Inventory

SECTION 313 CHEMICALS AND CHEMICAL CATEGORIES

- **Current list contains over 600 individual chemicals and chemical categories (See Table II of the EPA's TRI Reporting Forms and Instructions (RF&I) document).**
- **The list can change – check every year; changes listed in the front of the RF&I**

Toxics Release Inventory

THRESHOLD QUANTITY VS. RELEASE AND OTHER WASTE MANAGEMENT QUANTITIES

- **IF chemical activity quantity exceeds the threshold for manufacture, process, or otherwise use, you must report for this chemical, REGARDLESS of the amount of the chemical released or otherwise managed as waste (*i.e.*, recycling, treatment, and energy recovery).**

Toxics Release Inventory

- ***What is reported on the FORM R?***
  - **Part I**
    - **Facility Information**
      - **Facility name**
      - **Address**
      - **Certification**
      - **Technical and public contact**
      - **SIC code(s)**
      - **ID numbers**
      - **Parent company information**

Toxics Release Inventory

- **Part II of Form R:**
  - **Chemical identification, uses, and maximum quantity on-site during the reporting year**
  - **Quantities released on-site to air, land, water**
  - **Quantities transferred off-site for waste management**
  - **Description of on-site treatment, recycling, and energy recovery**
  - **Pollution Prevention Tracking:**
    - **Total quantities released and total quantities recycled, treated, or energy recovered for past year, current year, and upcoming years**
    - **Description of source reduction activities**

Toxics Release Inventory

An example of a portion of the Form R:

PART II. Section 5: Releases to Land On-Site

SECTION 5. QUANTITY OF THE TOXIC CHEMICAL ENTERING EACH ENVIRONMENTAL MEDIUM ONSITE (continued)				
		NA	**A. Total Release** (pounds/year*)(enter range code or estimate**)	**B. Basis of Estimate** (enter code)
5.4.1	Underground Injection onsite to Class I Wells	☐		
5.4.2	Underground Injection onsite to Class II-V Wells	☐		
5.5	Disposal to land onsite			
5.5.1A	RCRA Subtitle C landfills	☐		
5.5.1B	Other landfills	☐		
5.5.2	Land treatment/application farming	☐		
5.5.3A	RCRA Subtitle C surface impoundments	☐		
5.5.3B	Other surface impoundments	☐		
5.5.4	Other disposal	☐		

Toxics Release Inventory

HOW TO REPORT

- File a TRI report (a Form R or a Form A Certification Statement) for each Section 313 chemical exceeding an activity threshold
- Submit to U.S. EPA, and either designated state officials or designated tribal office by July 1st for preceding calendar year's activities
  - July 1, 2005 (January 1 - December 31, 2004 activities)
- Approximately 23,000 facilities file a total of 90,000 forms each year.

Toxics Release Inventory

What assistance is offered to help facilities determine if TRI reporting is required, and if reporting is required, what tools are available to help complete and submit the reporting forms?

Toxics Release Inventory

- **TRI WEBSITE**: www.epa.gov/tri
  - **Program Updates (*e.g.*, proposed rules)**
  - **Guidance Documents**
    - **Q & A**
    - **Industry Specific**
    - **Chemical Specific**
  - **Reporting Forms & Instructions**
  - **TRI workshops**
  - ***TRI-ME* Software**
  - **Regional and State TRI Contacts**
    - **Hotline: (703) 412-9810**

Toxics Release Inventory

TRI REPORTING SOFTWARE

- **The TRI-Made Easy *(TRI-ME)* Reporting Software is mailed to facilities with the Reporting Forms and Instructions package. *TRI-ME* can also be downloaded from www.epa.gov/tri.**
- **The *TRI-ME* software is an interactive, intelligent, user-friendly software program that assists facilities in determining and completing their TRI reporting obligations.**
- **Guides facilities in completing the Form R by explaining each element of the form through a questionnaire format.**

Toxics Release Inventory

Toxics Release Inventory

TRI-ME Tabs

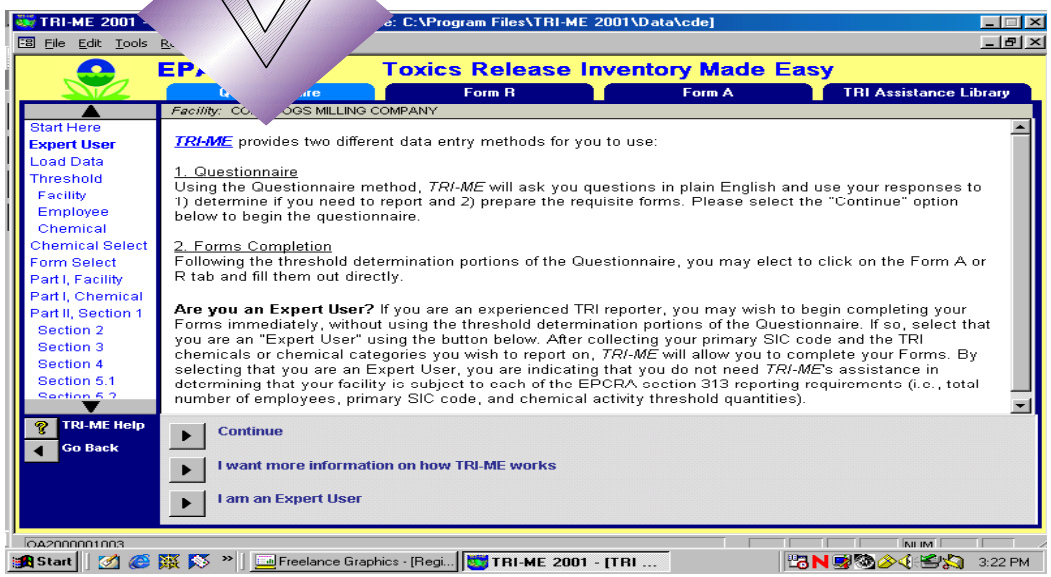
The TRI-ME Tabs at the top of the main portion of the screen allow you to move back and forth among the different parts of the program. When starting-up TRI-ME for the first time, you begin in the Questionnaire. The questionnaire uses a straightforward question-and-answer process to guide you through the reporting p[illegible]s. Clicking highlighted key terms and acronyms (displayed in blue text) in the Questionna[illegible]ll display a pop up box containing detailed definitions.

T
R
I

Toxics Release Inventory

T
R
I

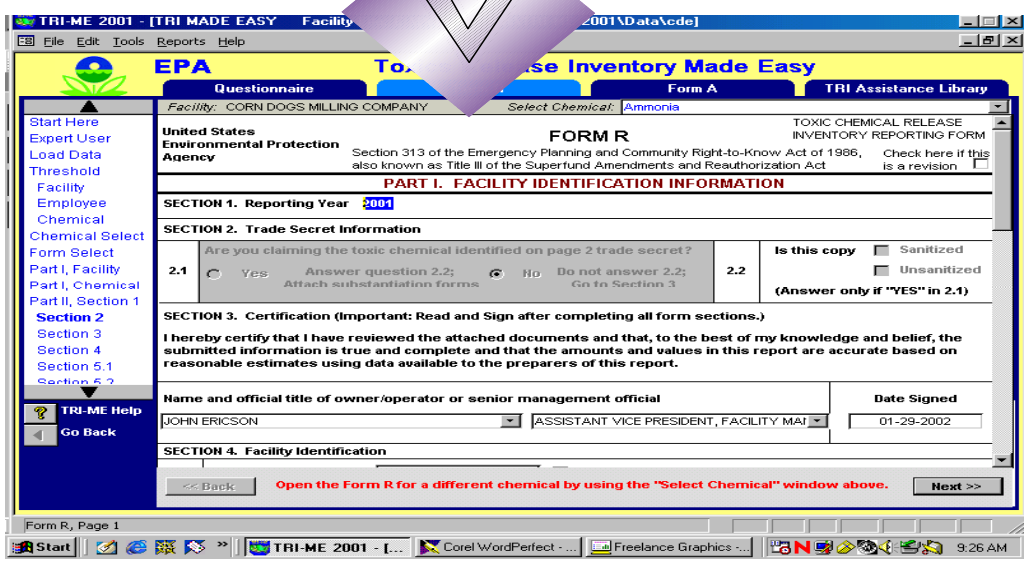
Questionnaire- *The TRI-ME questionnaire walks the user through the analysis to determine if their facili[illegible]quired to file TRI forms for listed toxic chemicals.*



Toxics Release Inventory

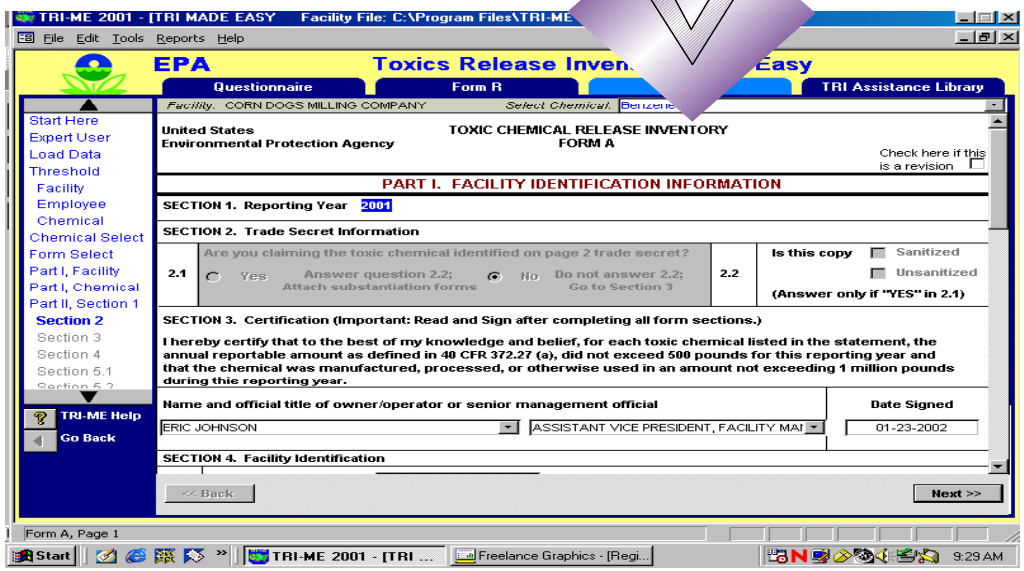
T
R
I

Form R Tab – *Allows the user to select which reporting form – Form R or Form A – a facility wishes to submit for each toxic chemi… he user does not make a selection – TRI-ME defaults to Form R for all toxic chemicals.*



Toxics Release Inventory

Form A Tab - *Allows the user to select which reporting form - Form R or Form A - a facility wishes to submit for each toxic chemical. If the user does not make a selection - TRI-ME defaults to Form R for all toxic chemicals.*



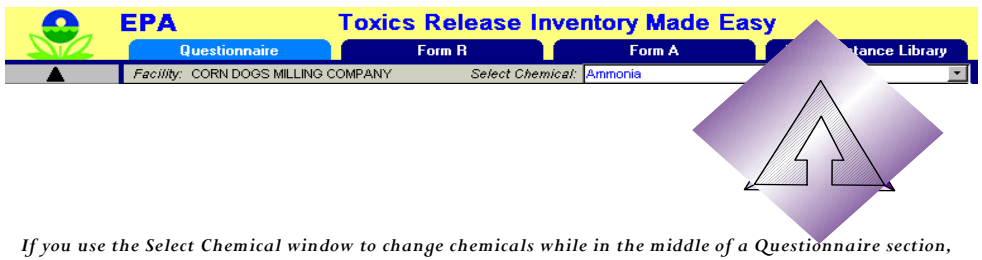
T
R
I

Toxics Release Inventory

T
R
I

Select Chemical Window

The Select Chemical window enables you to select the TRI chemical that you wish to work with. This select chemical window is only displayed in those places where chemical-specific Form R or Form A information is being collected. The list of chemicals provided in the window includes all Form R chemicals and all Form A chemicals.



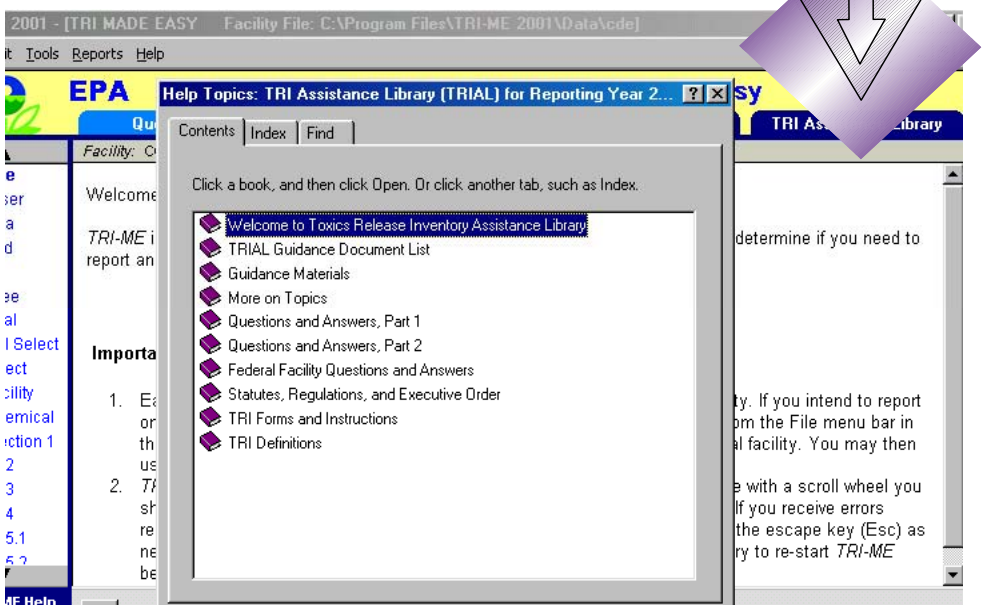
If you use the Select Chemical window to change chemicals while in the middle of a Questionnaire section, TRI-ME will return you to the beginning of that section for the new chemical to ensure that you are given all applicable guidance.

If you use the Select Chemical window to change chemicals while in using the Form R or Form A Tab, TRI-ME will open the Form associated with the new chemical, but remain in the same Form section.

Toxics Release Inventory

T
R
I

TRI Assistance Library – *Is a searchable, indexed file that contains the statutes, the regulations and the key TRI guidance documents a facility is likely [illegible]eed for TRI reporitng.*

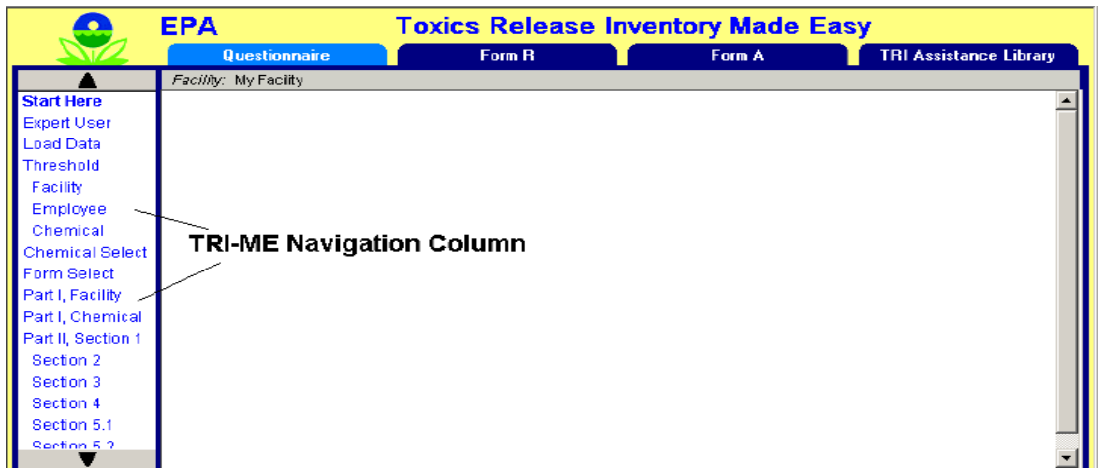


Toxics Release Inventory

T
R
I

TRI-ME Navigation Column

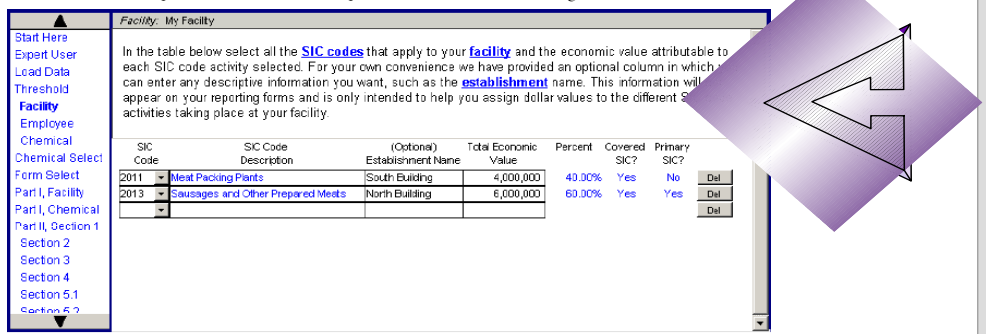
The Navigation Column allows you to quickly move to the first screen of any part of the Questionnaire that is available. Those items displayed in blue text are areas of the Questionnaire that you may access.



Toxics Release Inventory

TRI-ME Screen Window

When working in the Questionnaire, this part of the TRI-ME screen contains informational text combined with questions that guide you through the process of completing your threshold calculations or your Forms. In some cases, the screen is split into text and a simple form used for collecting data as shown below.



In other cases, data is collected on the basis of your selection of a "Continue" Button based on questions posed in the text.

When working in the Form R or Form A Tab, this part of the TRI-ME screen contains a facsimile of the Form R or Form A and enables you to enter information directly into the facsimile forms. The applicable Form R or Form A page number is displayed in the lower left-hand corner of the TRI-ME screen.

T
R
I

Toxics Release Inventory

T
R
I

TRI-ME "Continue" and More on [Topic] Buttons

TRI-ME "Continue" Buttons

On each TRI-ME Questionnaire screen there are "Continue" Buttons located in the bottom portion of the TRI-ME screen. Use the TRI-ME continue buttons to answer questions posed in the questionnaire text and to navigate throughout the system.

TRI-ME More on [Topic] Buttons

On certain TRI-ME Questionnaire screens there are More on [Topic] Buttons that open help topics contained within the TRI Assistance Library. These help topics are provided to assist you in answering questions or providing data for that specific portion of the Questionnaire. These help topics can be accessed directly by clicking the TRI Assistance Library tab and double-clicking on the "More on Topics" under the TRIAL Contents tab.

Toxics Release Inventory

TRI-ME REPORTING SOFTWARE

- **Prevents facilities from making common errors while completing the Form R.**
- **Checks (validates) the forms to identify critical errors that must be corrected before submitting the forms to EPA. Also, suggests potential errors for user review.**

Toxics Release Inventory

***TRI-ME* REPORTING SOFTWARE**

- **Allows electronic Internet submission of TRI forms via the EPA's Central Data Exchange (CDX), which is the point of entry on the Environmental Information Exchange Network (Exchange Network) for environmental data exchanges to the Agency.**
- **Security – Your information is protected by username and password (and secret question/answer) that you create.**

Toxics Release Inventory

Benefits of Submitting via *TRI-ME* and CDX

- **Significantly reduces data errors by using *TRI-ME* and eliminating manual data capture**
- **Electronic Signature allows for *Paperless Filing***
- **Facilities receive instant email confirmation receipt**
- **CDX submissions are processed faster than paper submissions, which leads to faster Facility Data Profile (FDP) access. This means reduced data collection costs for EPA, States, and the regulated community.**

Toxics Release Inventory

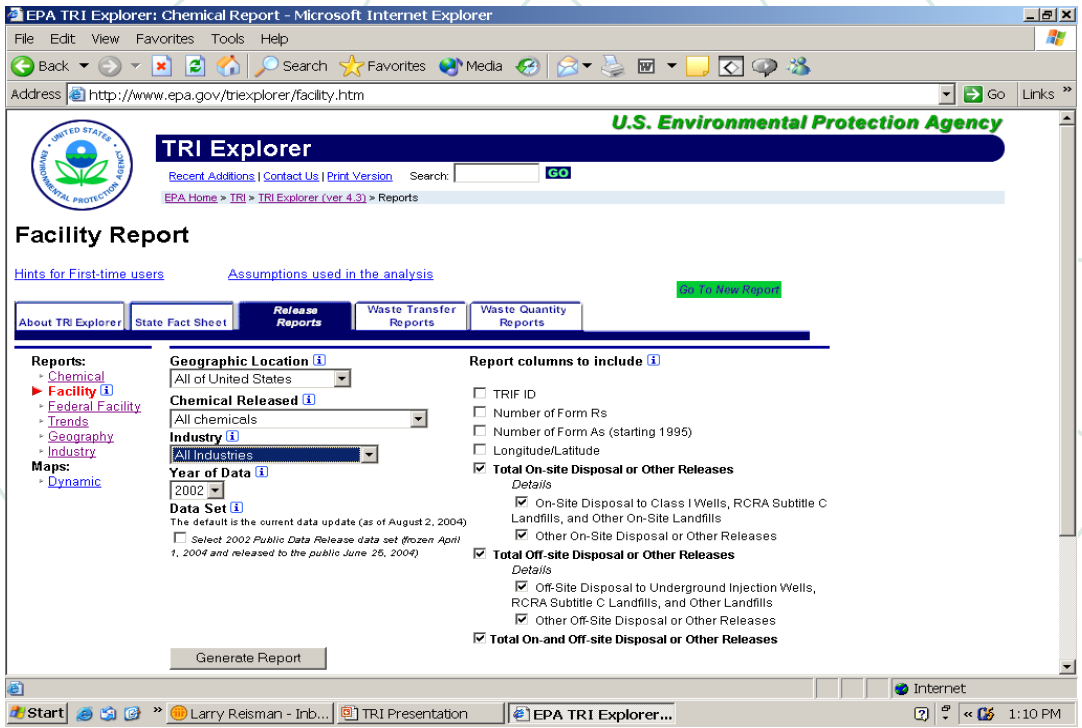
Accessing and Using the TRI Data

Toxics Release Inventory

Data Access

TRI Explorer provides fast and easy access to the TRI data:

http://www.epa.gov/triexplorer/



Envirofacts Tables used in TRI Explorer

- Envirofacts data is in ORACLE in a normalized DBMS
- The DBMS is ideal for storage but not for real time Web application reporting
- The Web application would require very complicated SQL queries that would take a long time to execute
- The SQL queries do not change over time
- The solution is to create de-normalized tables that are optimized for the TRI Explorer Web application

TRI Explorer

TRI Data and Web Applications

List of TRI tables in Envirofacts

TABLE_NAME

TRI_CHEM_ACTIVITY
TRI_CHEM_INFO
TRI_CODE_DESC
TRI_COUNTY
TRI_ENERGY_RECOVERY
TRI_FACILITY
TRI_FACILITY_DB
TRI_FACILITY_DB_HISTORY
TRI_FACILITY_HISTORY
TRI_FACILITY_NPDES
TRI_FACILITY_NPDES_HISTORY
TRI_FACILITY_RCRA
TRI_FACILITY_RCRA_HISTORY
TRI_FACILITY_SIC
TRI_FACILITY_SIC_HISTORY
TRI_FACILITY_UIC
TRI_FACILITY_UIC_HISTORY

TABLE_NAME

TRI_OFF_SITE_TRANSFER_LOCATION
TRI_ONSITE_WASTESTREAM
TRI_ONSITE_WASTE_TREATMENT_MET
TRI_POTW_LOCATION
TRI_RECYCLING_PROCESS
TRI_RELEASE_QTY
TRI_REPORTING_FORM
TRI_SOURCE_REDUCT_METHOD
TRI_SOURCE_REDUCT_QTY
TRI_SUBMISSION_SIC
TRI_TABLE_ID_NAME
TRI_TRANSFER_QTY
TRI_WATER_STREAM
TRI_ZIP_CODE

TRI Explorer

TRI Data and Web Applications

- Typically the tables in the DBMS have relatively few attributes but many tuples spanning across multiple years
- The chemical amounts are reported to EPA in pounds (or grams) or as a range code, and are stored in Envirofacts as such. However, they are reported in TRI Explorer only in pounds (or grams)
- When an estimate is reported, there will be a letter code that identifies the method that applies to the largest portion of the total estimated release quantity

TRI Explorer

Creating SAS Datasets

Why SAS?

- Back in 1998 when the first version of TRI Explorer was being developed, SAS had the capability to run Web services which executed compiled macros and read tables in the back-end and created html formatted reports that were written to Web browsers via the common gateway interface. The ability to create customizable html reports was a big advantage over competing products.
- SAS offers tools that can be used for analysis with minimal additional effort

TRI Explorer

Creating SAS Datasets

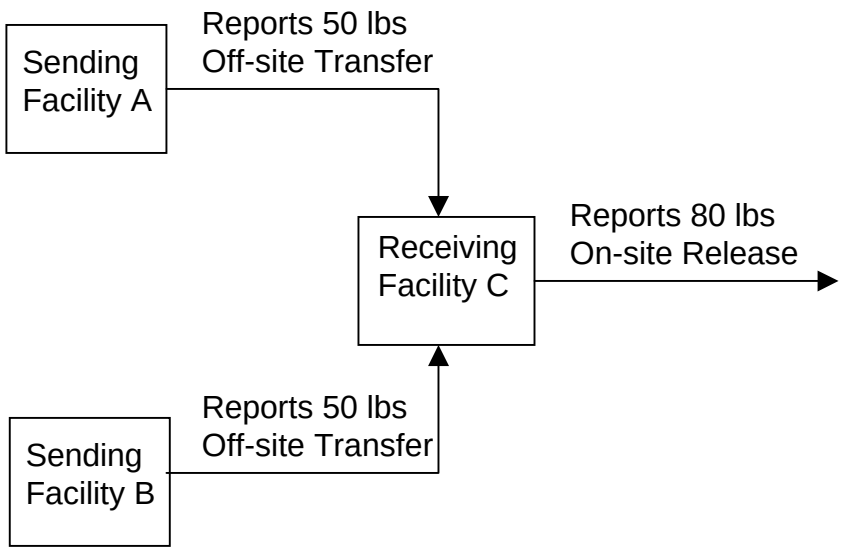
Extracting Data from Envirofacts

- We use a set of SAS macro's to read the ORACLE tables and create denormalized SAS tables. The process involves selecting only EPCRA section 313 chemicals, creating chemical groups, converting all release amounts to pounds, creating 2 digit industry codes and selecting the appropriate 'M' code when a chemical is transferred off-site
- When a facility receives chemicals from one or more facilities then reports the sum of those chemicals as releases, we apply only the portion reported by the receiving facility

TRI Explorer

Creating SAS Datasets

Example



The omitted amounts applied to facilities A and B are 40 lbs

TRI Explorer

Creating SAS Datasets

- The end result after running the processes just mentioned is a set of denormalized flat files, one per year. The TRI program started collecting data in 1987 but TRI Explorer starts reporting from 1988
- Each tuple in the flat file contains a complete FORM R or FORM A
- The primary key is the Document Control Number

TRI Explorer

Creating SAS Datasets

To optimize the performance of the Web application we create 4 sets of pre-summarized tables. Each set is composed of a collection of tables - one for each year since 1988

The 4 sets of summary tables are by

- county and chemical
- state and chemical
- US and chemical
- Facility and chemical

All the tables make up a star schema but we did not set it up as a data warehouse. The flat files described in the previous slide are the fact tables in the star schema

TRI Explorer

Web Services

TRI Explorer is a highly valued tool in EPA because it is easy to use and it gives the user a panoramic view of the information. A user can generate a useful report in two or three mouse clicks, unlike some applications which have a learning curve

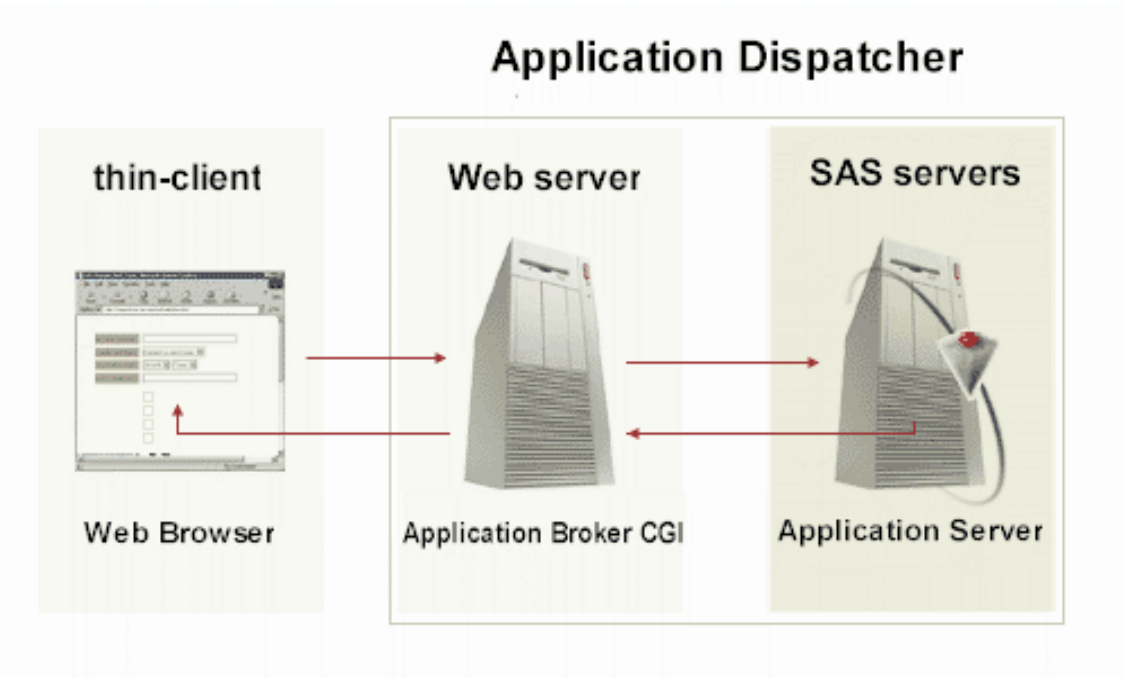
TRI Explorer's Web services use SAS/IntrNet software, a product from SAS. You can get more details about it from http://support.sas.com/rnd/web/intrnet/index.html

Web services have 3 components;

- User interface, usually a web browser
- Web server, www.epa.gov
- Back end database server

TRI Explorer

Web Services



TRI Explorer uses SAS/IntrNet's Application Dispatcher which is composed of two parts;

- Web server
- SAS server

TRI Explorer

Web Services

Web Server

- The Web server is installed in the machine that hosts the Web site. It is a common gateway interface (CGI) program that receives request from Web browsers and other front end tools and directs them to the back end SAS server via specific TCP ports.
- Requests come in the form of a uniform resource locator (URL) with a specific syntax. The syntax in the URL identifies which SAS server to invoke and the task that it will perform.
- The Web server, also known as the application broker, is configured once during installation and runs continuously – as long as the Internet host machine is serving Web pages.

SAS server

- In a production environment, it is always advisable to run the SAS server on a separate host machine that is preferably inside a firewall to protect it and to optimize on performance
- The SAS server is a process that waits for request from the application broker on specific pre-defined ports
- There are usually multiple SAS servers, also known as application servers, all waiting for requests from a broker.
- Upon receiving a request, the application server establishes a session with the application broker for sending and receiving data. Sessions are typically set to expire after 3 minutes, and are configurable in the application broker
- The application server can be configured to spawn multiple process to handle simultaneous requests

TRI Explorer

Web Services

SAS server

- Each SAS server can be configured independently and is capable of running native SAS commands or compiled SAS procedural macros
- In TRI Explorer we compile the SAS macros. There are 48 macros, with a combined total of more than 20,000 lines of code
- Each macro is designed to perform a specific set of tasks with only a minimal amount of overlap
- The SAS server can be configured to log all or some of the code being executed. The log files will add an extra burden on the application but they are useful when maintaining a large application

TRI Explorer

Web Services

SAS server process flow

- All processing in TRI Explorer is flows through a single point of entry where global variables and libraries are defined
- Control is passed to other macros based on the options selected by the user. SAS data tables are read and processed at this level and the report is initiated
- After the body of the report is populated with data, control is passed to other macros that create footnotes, where applicable
- The report in the form of an HTML is then routed via the application server, back through the open session to the application broker which in turn sends it to the user's browser
- In some cases when the user has a slow browser connection and/or a slow application server connection the session will time out and no report will be generated. SAS will automatically send an appropriate message indicating so.

TRI Explorer user interface

- The user interface is normally a Web browser, but some Web developers bypass the browser send URL requests directly to the application broker
- The selects options in HTML forms, most are drop down lists but some are radio buttons or check boxes. Other option trigger pop up windows. After making the selections the user generates a report by clicking on the generate report button
- There is embedded JavaScript code in the HTML files and more JavaScript script invoked from the server to control the selections made in the HTML forms
- There is a Perl script in the Web server's CGI that is used to redraw HTML pages when drilling down from some options

TRI Explorer

Web Services

TRI Explorer reports

- Most reports in TRI Explorer will not fit on a standard screen and we give the user an option to pick only those columns of data they would like to see
- Some reports are in the form of a map. The map is a Java applet that is downloaded to the user. The map does require that a Java runtime module be installed as part of the browser.
- Most reports are downloadable to the user's computer as comma separated value files which they can then import to a spreadsheet for further analysis or printing

TRI Explorer

Any Questions?

Envirofacts

July 2005

OFFICE OF ENVIRONMENTAL INFORMATION

Overview

- Purpose
  - To provide an overview of EPA's Envirofacts Public Access System
- Agenda
  - Overview
  - Building the Database
  - The User Interface
  - System Management
  - Current Usage
  - Future Modernization

Background

OFFICE OF ENVIRONMENTAL INFORMATION

EPA's Historical Approach to Environmental Protection

- Media-specific legislation and regulations
- Regulatory focus
- Traditional business cycle
  - Science
  - Regulations
  - Enforcement

OFFICE OF ENVIRONMENTAL INFORMATION

Impact on Information Technology

- Media-specific information systems
- Lack of information sharing/integration
- Lack of ability to cross-manage systems
- Higher costs

EPA's IT History

- Modern infrastructure
- Good connectivity
- Challenges with data management

OFFICE OF ENVIRONMENTAL INFORMATION

EPA's Evolving Mission

- Cross-program integration
- Pollution prevention
- Geographic initiatives
- Voluntary compliance
- Community-based environmental protection

OFFICE OF ENVIRONMENTAL INFORMATION

Reinventing Environmental Information

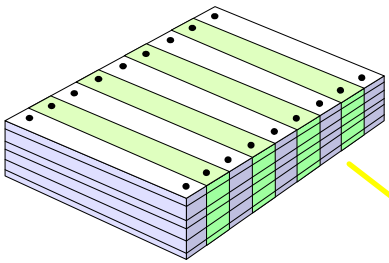
"We can create a seamless environmental information system that will minimize costs…and enhance access for the public and regulators."

Carol Browner
EPA Administrator
July 21, 1997

OFFICE OF ENVIRONMENTAL INFORMATION

REI Vision

Information Collection



Integrated Reporting
Electronic Data Collection
Burden Reduction

Information Integration

Data Standards
Data Registry
Data Warehouse
Reengineering
Programmatic Data

Information Access

Environmental Monitoring
Center for Environmental
Information & Statistics
Envirofacts
1-800
Community Right-to-Know

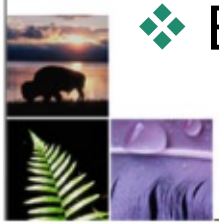
OFFICE OF ENVIRONMENTAL INFORMATION

How It Came About

OFFICE OF ENVIRONMENTAL INFORMATION

Timeline

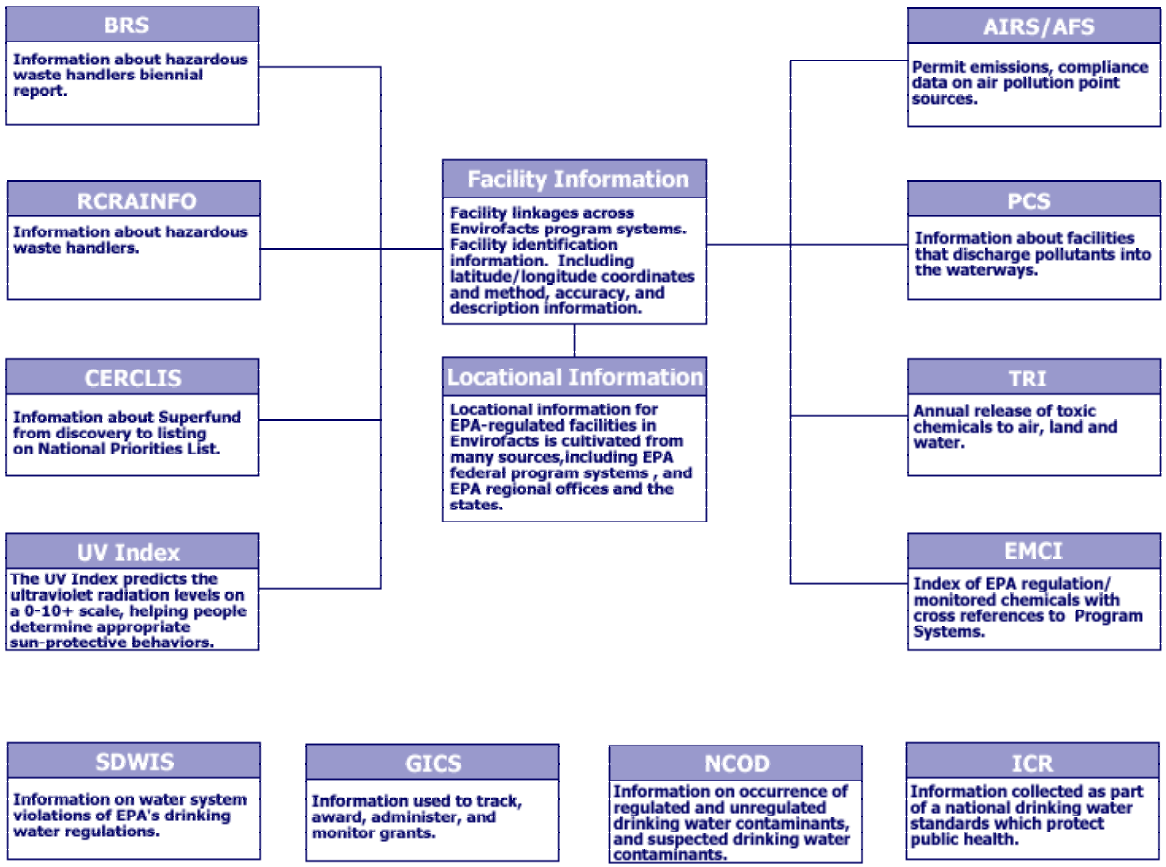
- Gateway “Universal” Interface – 1991
- Envirofacts Database with Facility Integration – 1992
- Earth Day – 1993
- Decision to go “Web-based” – 1994
- Live on the Internet – 1995
- Multisystem Query - 1995
- Chemical Integration - 1996
- Locational Data Improvement – 1996
- Enviromapper - 1998
- Facility Registry System - 1999
- ETL/Business Intelligence Modernization - 2006



Building the Database

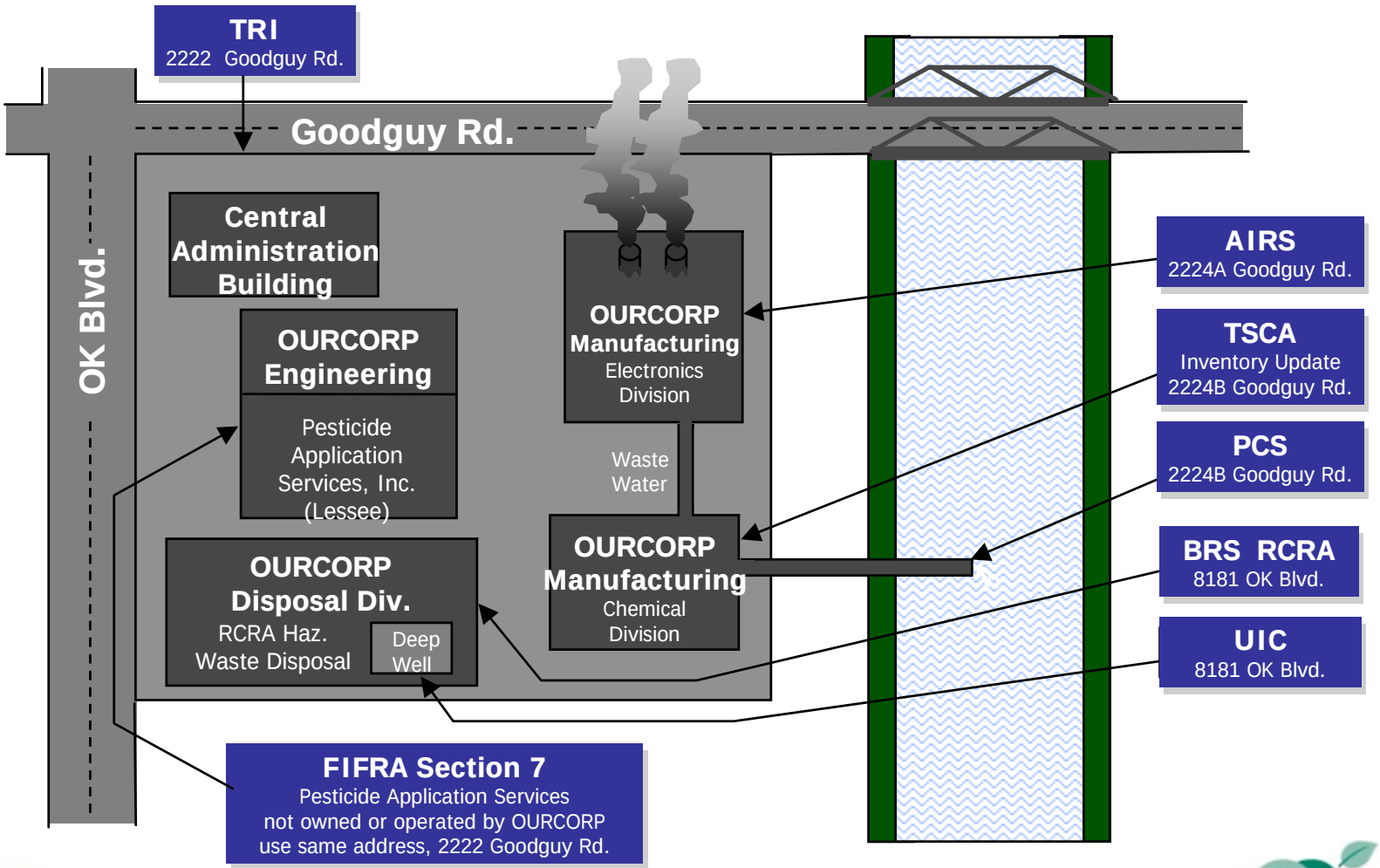
OFFICE OF ENVIRONMENTAL INFORMATION

Envirofacts Data Model



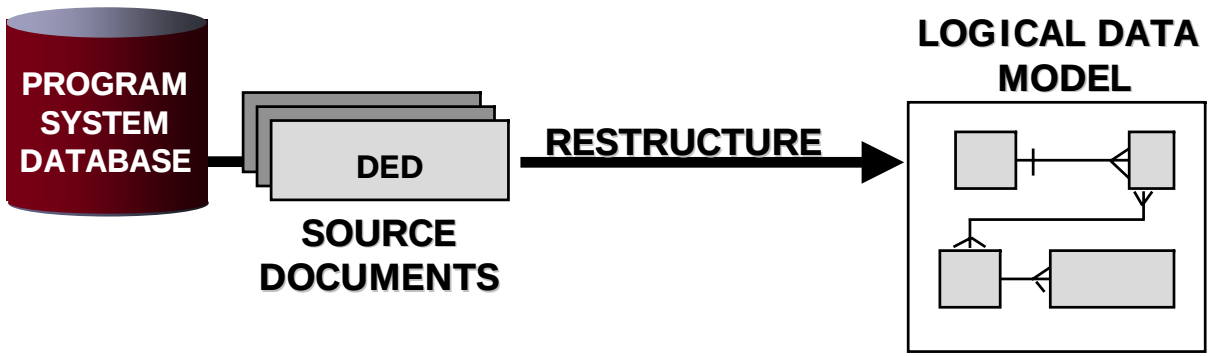
OFFICE OF ENVIRONMENTAL INFORMATION

OURCORP Facility Complex



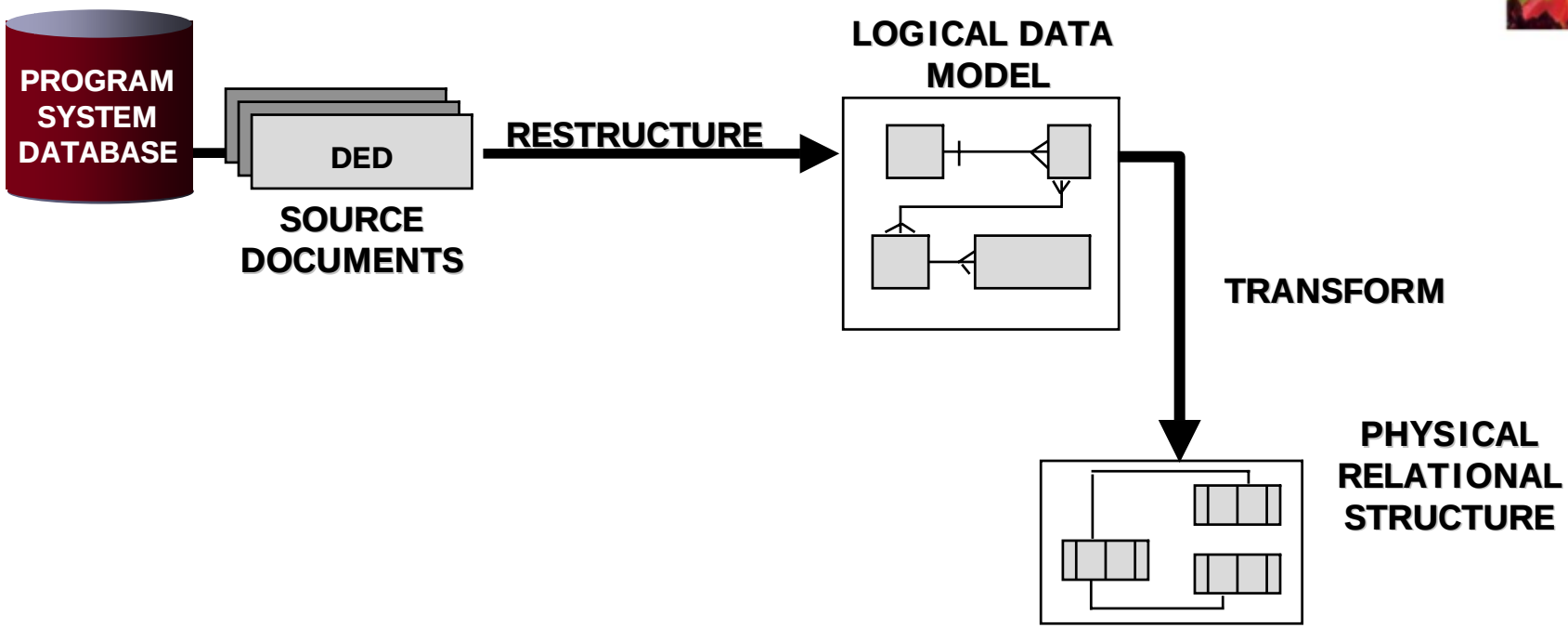
OFFICE OF ENVIRONMENTAL INFORMATION

Getting the data…



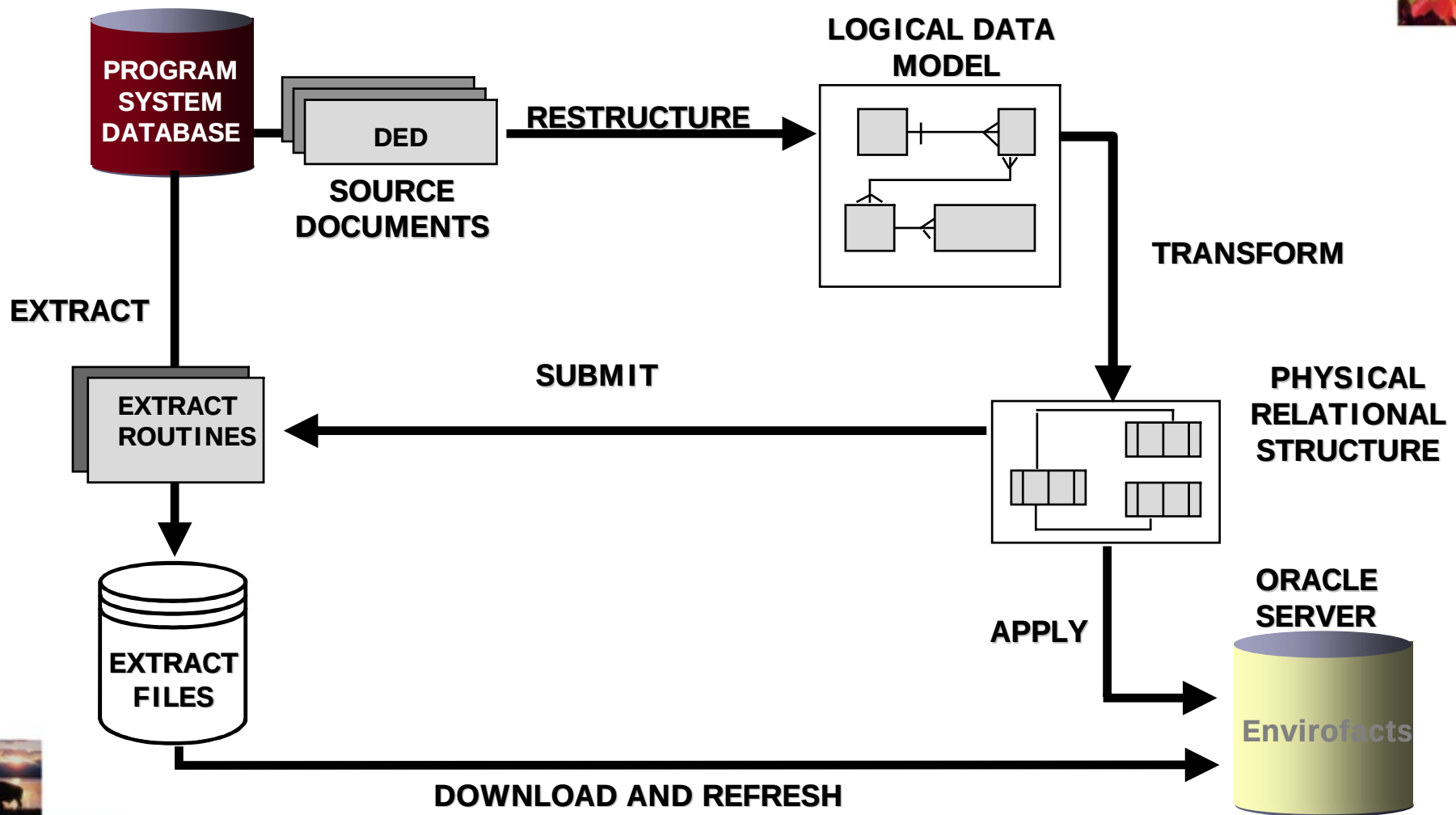
OFFICE OF ENVIRONMENTAL INFORMATION

Getting the data…



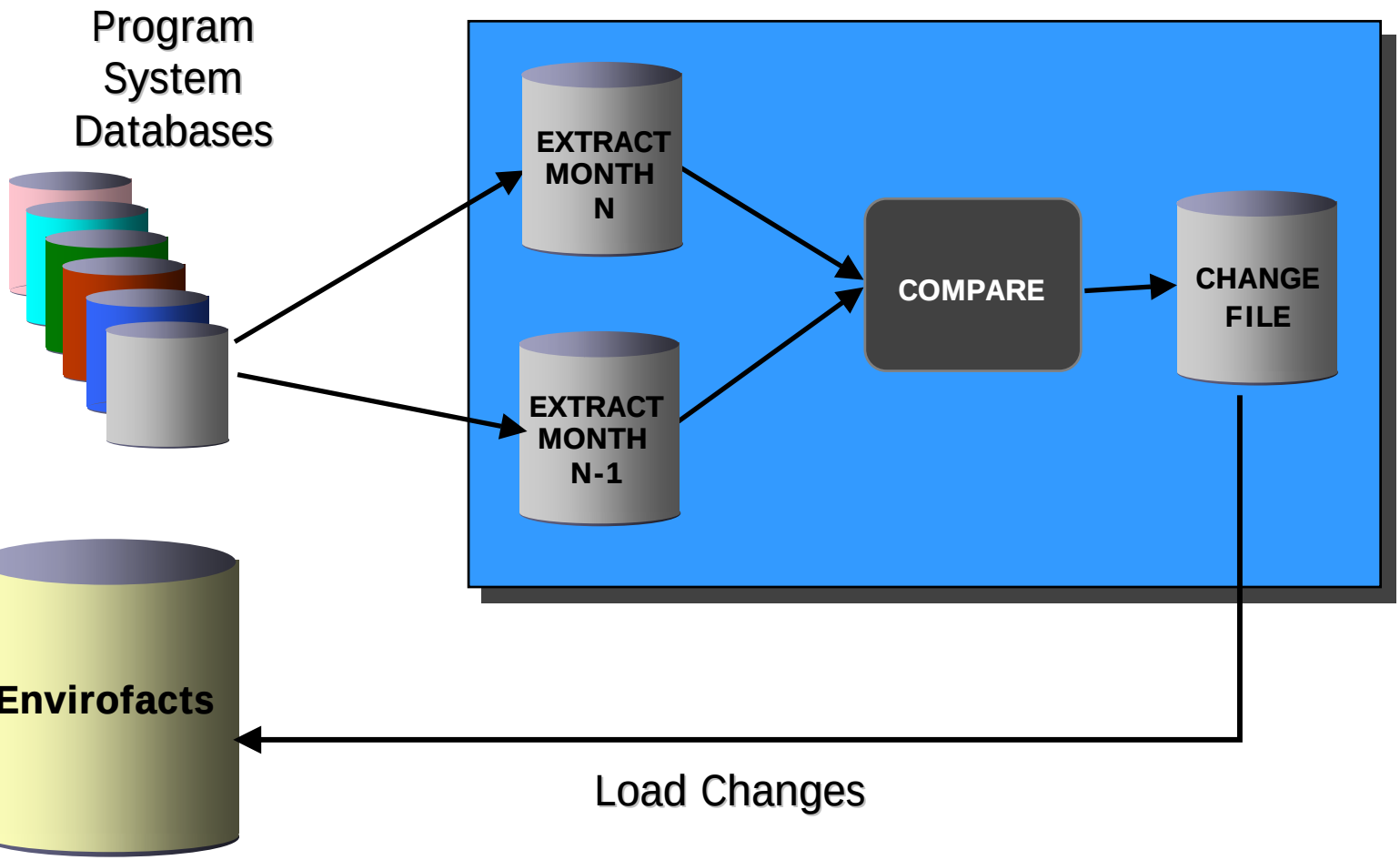
OFFICE OF ENVIRONMENTAL INFORMATION

Getting the data…



OFFICE OF ENVIRONMENTAL INFORMATION

Refresh Process



OFFICE OF ENVIRONMENTAL INFORMATION

Usage Statistics

- 90K Plus Distinct Hosts Served
- Avg. 1.5 Million Requests Per Month
- Avg. 500K Page Requests Per Day

OFFICE OF ENVIRONMENTAL INFORMATION

Live Demonstration

OFFICE OF ENVIRONMENTAL INFORMATION

Contact Information:

Bill Muldrow: muldrow.bill@epa.gov

OFFICE OF ENVIRONMENTAL INFORMATION

Data Marts/Data Warehousing

Environmental Information Systems Management Study Tour

July 5-13, 2005

OFFICE OF ENVIRONMENTAL INFORMATION

Agenda

- Data Warehouse Automation
- Data Bus Architecture
- ETL Tool Evaluation
- Data Quality Automation
- Automating the User's View

OFFICE OF ENVIRONMENTAL INFORMATION

Data Warehouse Automation

OFFICE OF ENVIRONMENTAL INFORMATION

Envirofacts Data Warehouse

SOURCE SYSTEMS

MAINFRAMES
- AFS
- PCS
- GICS
- SIDWIS

TEXTFILES
- CERCLIS
- ERAMS
- BF/BMS
- TRIS

DUMP FILES
- RCRIS

PROCESS ENGINE

SQL*Loader
- Load Destination Tables

PERL
- Validity Check
- Formatting
- Renaming Source Files
- Truncate Destination Tables

Oracle Import/Export Utility
- Load Data
- Extract data

Manual Data Validation

EnviroFacts Data Warehouse

OFFICE OF ENVIRONMENTAL INFORMATION

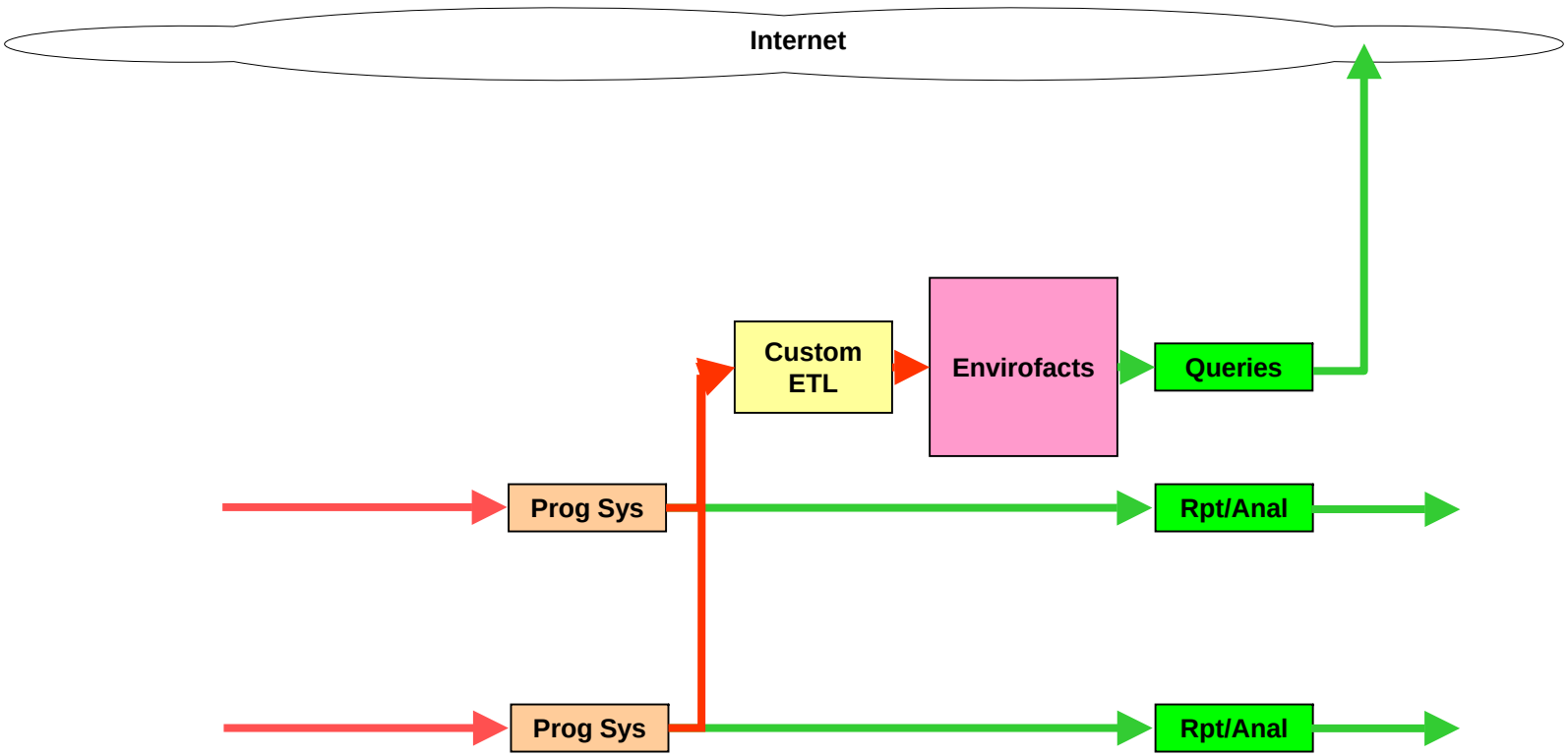
Why Environmental Data Marts?

- Support Environmental Analysts
  - Science Advisory Board request for timely, consistent, actionable access for researchers
  - Increases value of air quality monitoring network by making existing data more accessible to users
- Get IT out of the expensive query and reporting business
  - Provide a more simplistic, higher quality reporting engine
  - Provide alternative to extensive customization
- Improve Access Security
  - Control who is entitled to access and update data
  - Prevent unauthorized or illegal access to data

Historically…

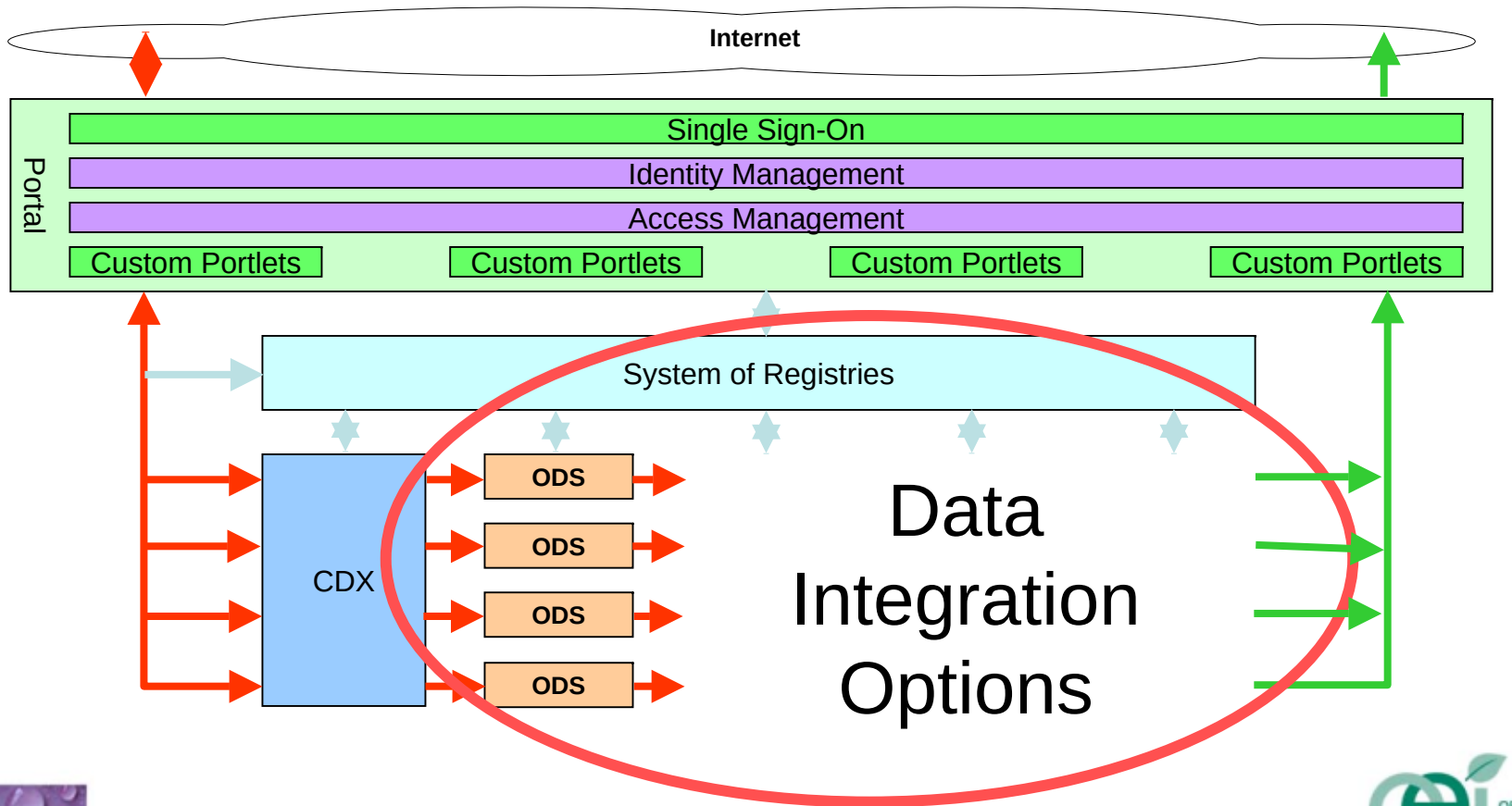
OFFICE OF ENVIRONMENTAL INFORMATION

Current data integration…



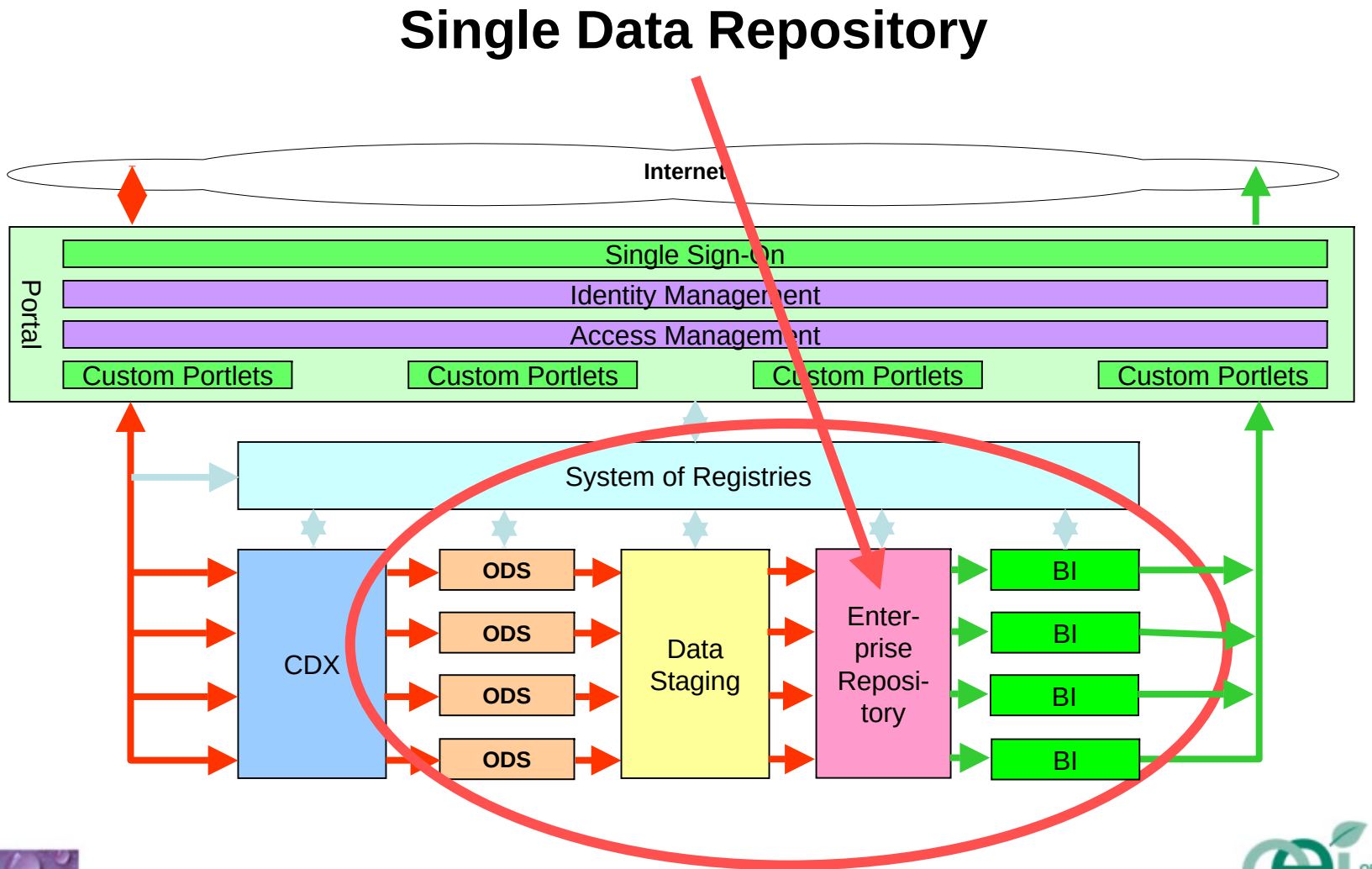
OFFICE OF ENVIRONMENTAL INFORMATION

Comprehensive Target Architecture



OFFICE OF ENVIRONMENTAL INFORMATION

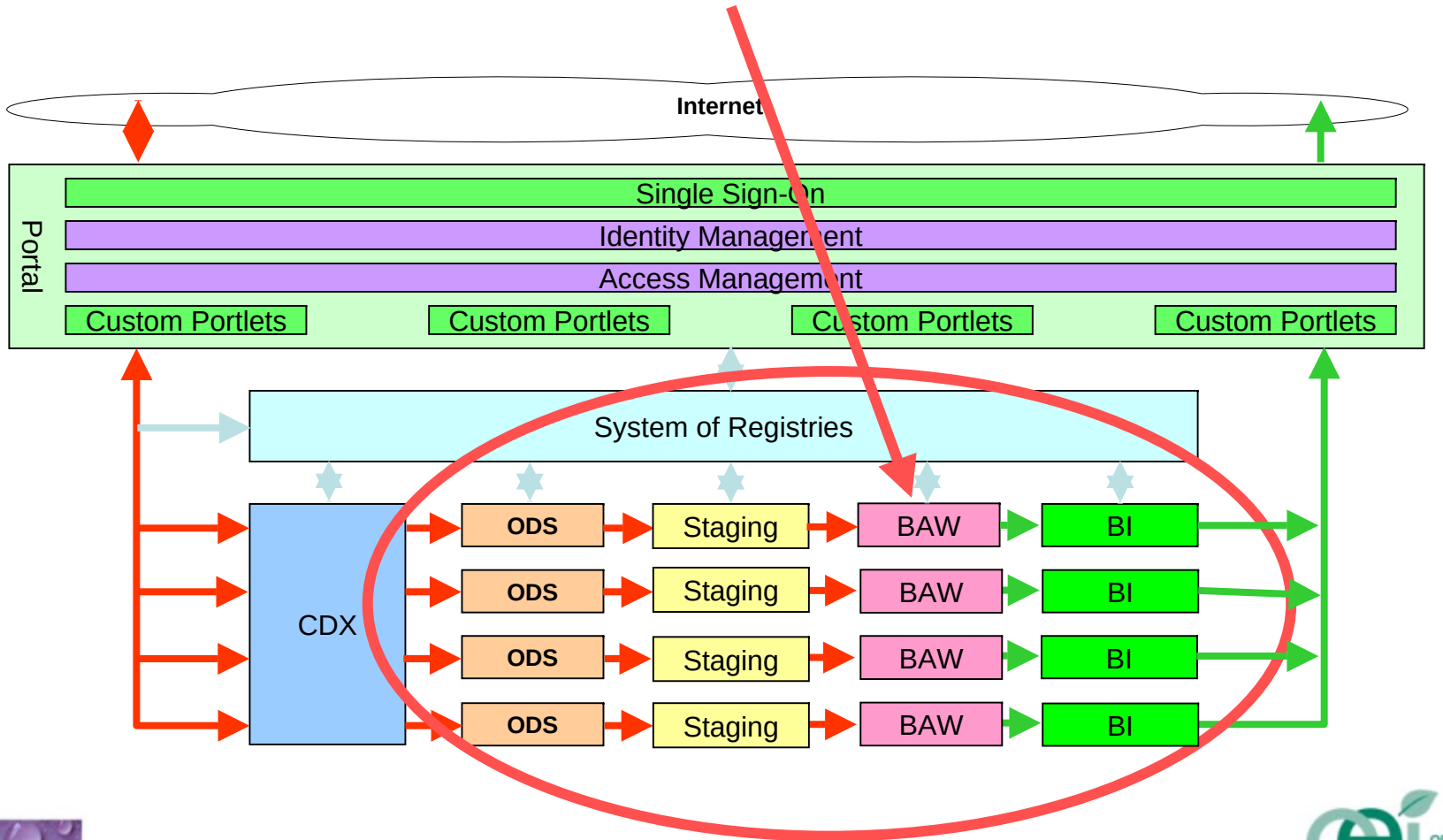
Alternatives…



OFFICE OF ENVIRONMENTAL INFORMATION

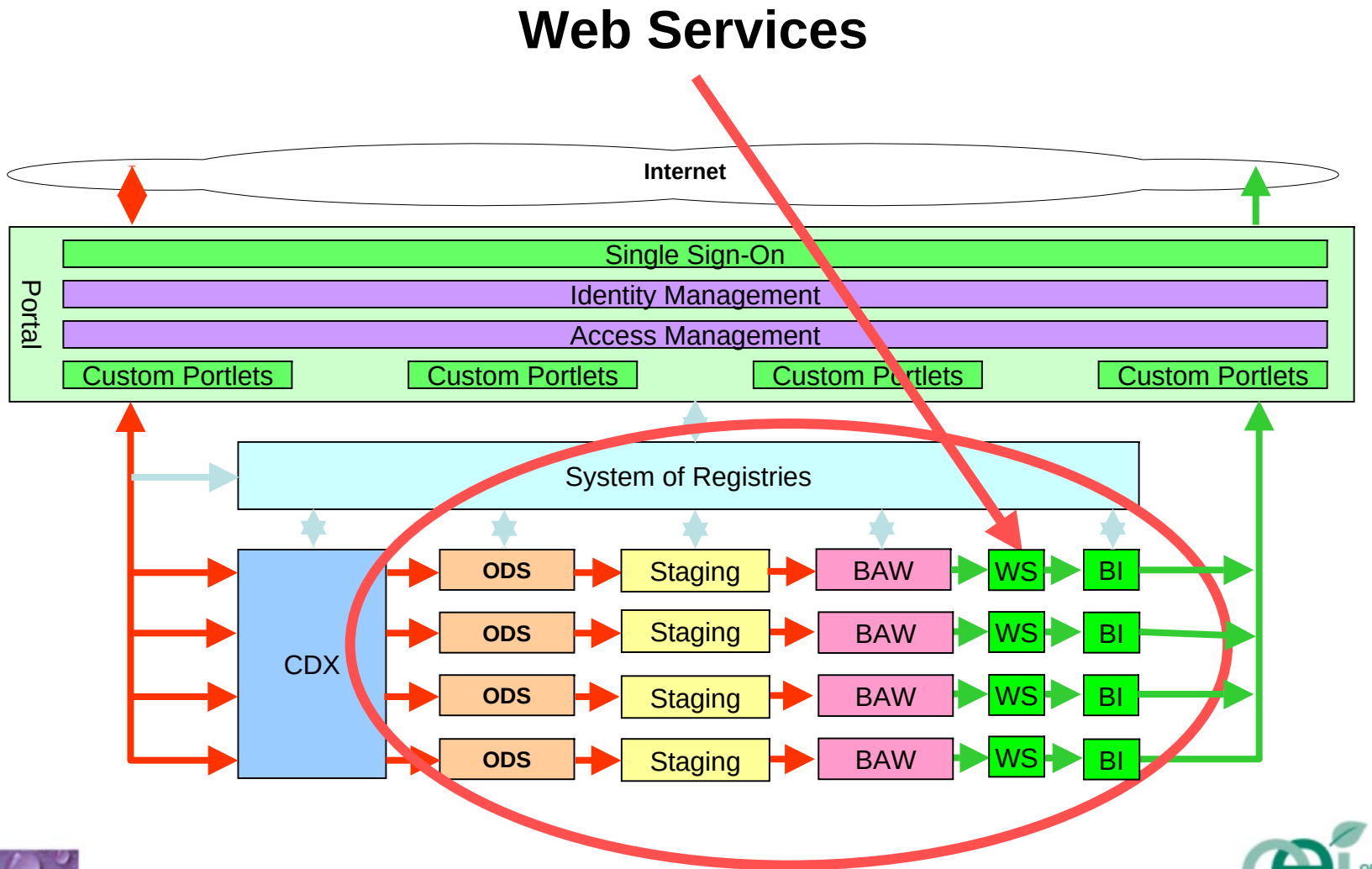
Alternatives…

Framework of Business Area Warehouses



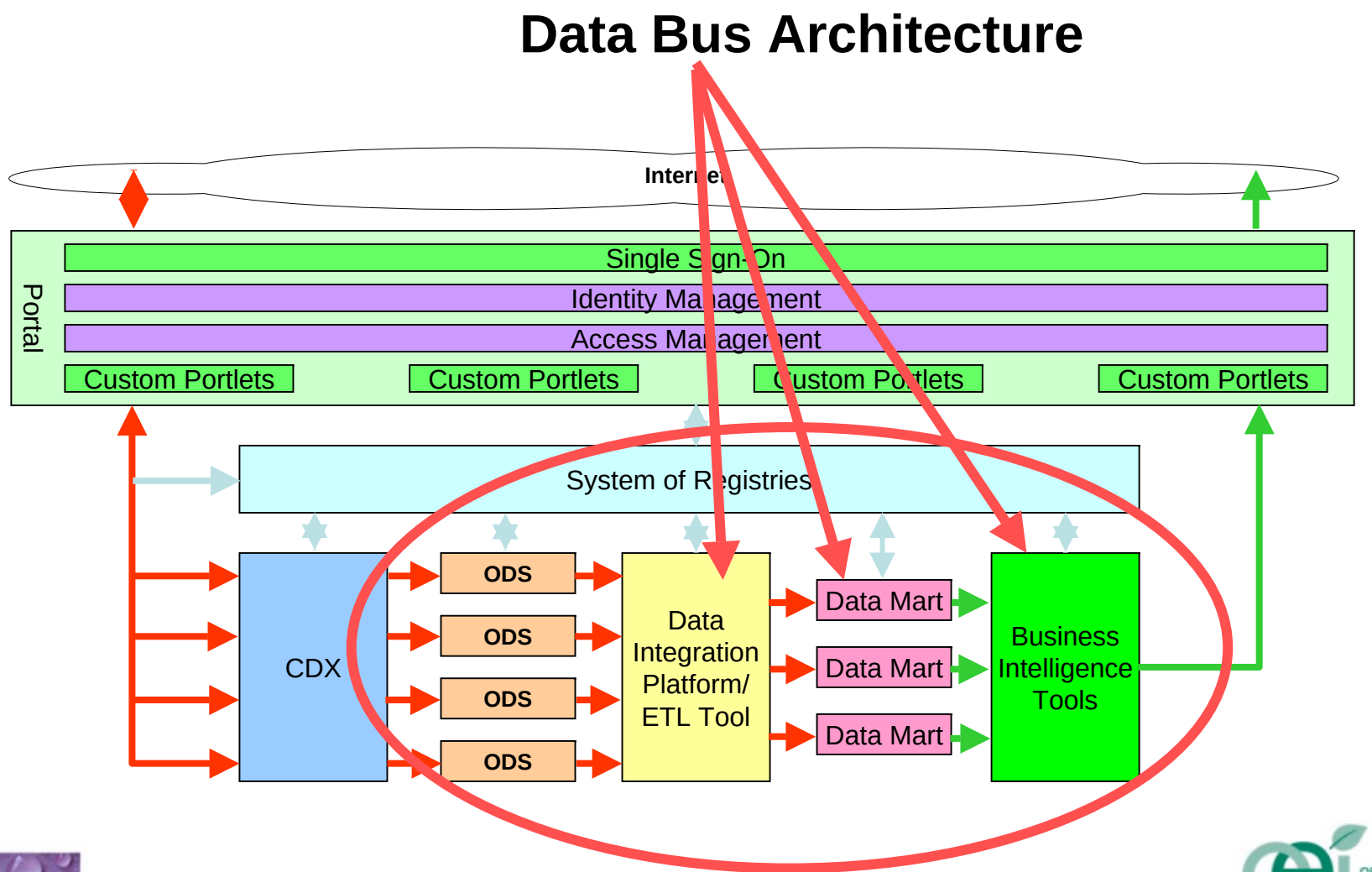
OFFICE OF ENVIRONMENTAL INFORMATION

Alternatives…



OFFICE OF ENVIRONMENTAL INFORMATION

Data Integration Target Architecture

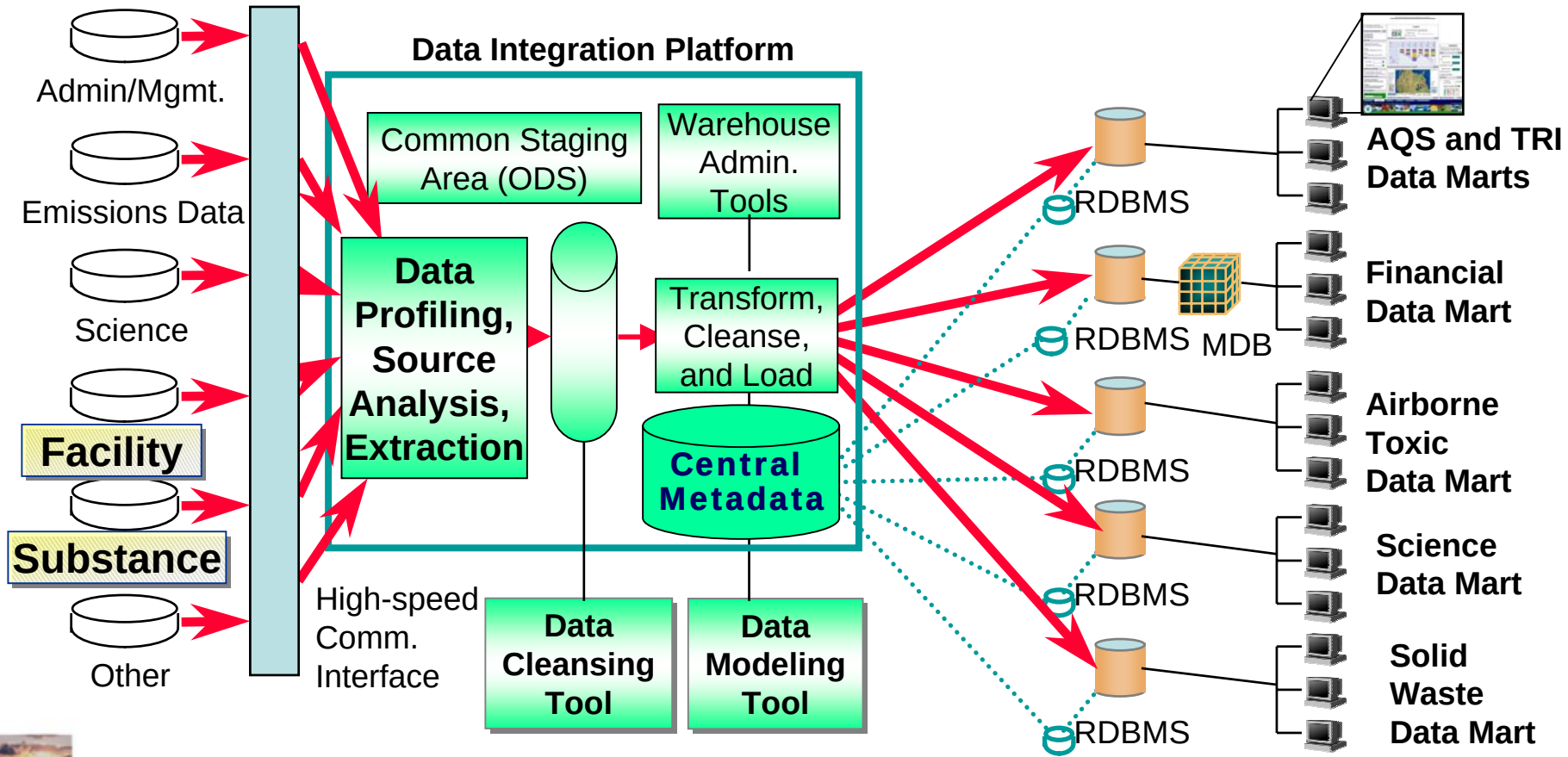


OFFICE OF ENVIRONMENTAL INFORMATION

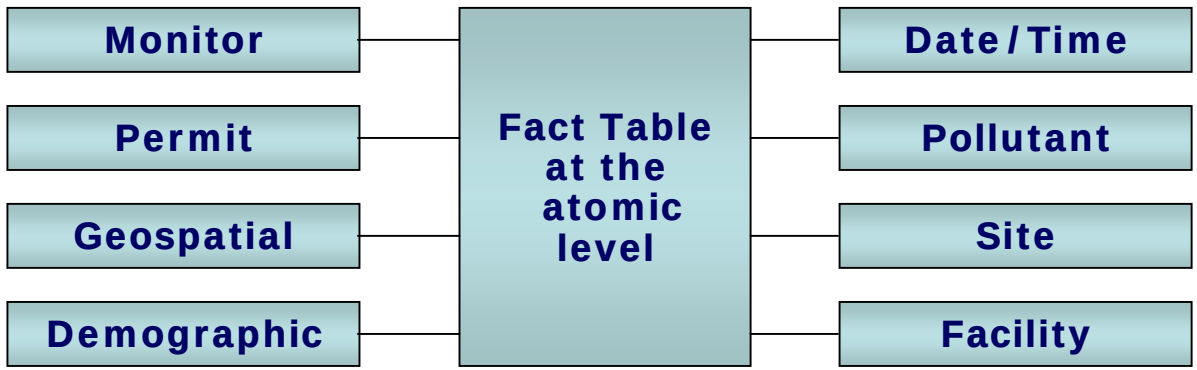
Data Bus Architecture

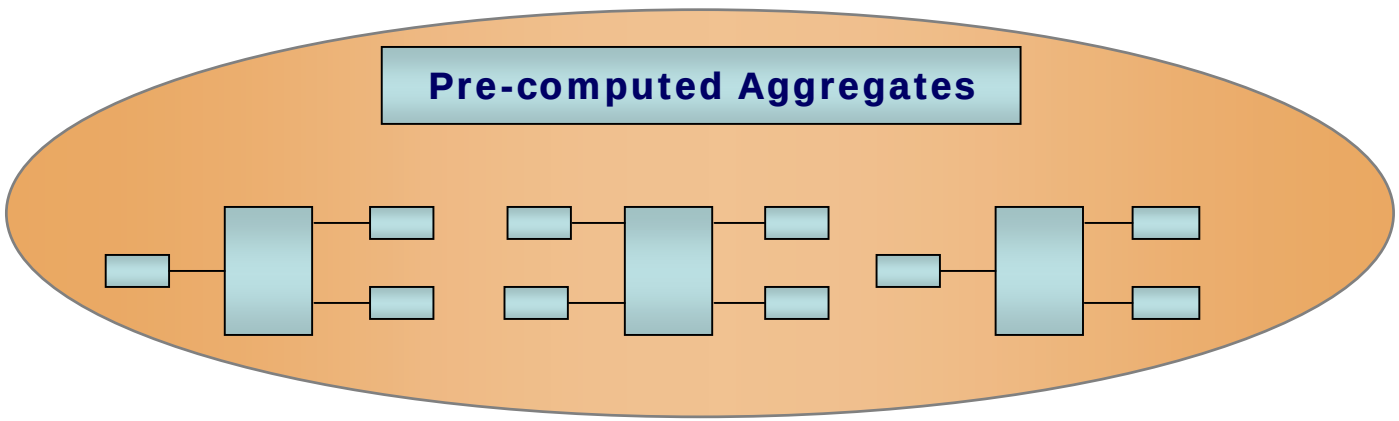
OFFICE OF ENVIRONMENTAL INFORMATION

Data Bus Architecture



Target Data Model
Star Schema and Aggregates





Facts are Navigated by One Step From STAR Point to Fact
(Vs. Complex, multi-join query in warehouse approach)

OFFICE OF ENVIRONMENTAL INFORMATION

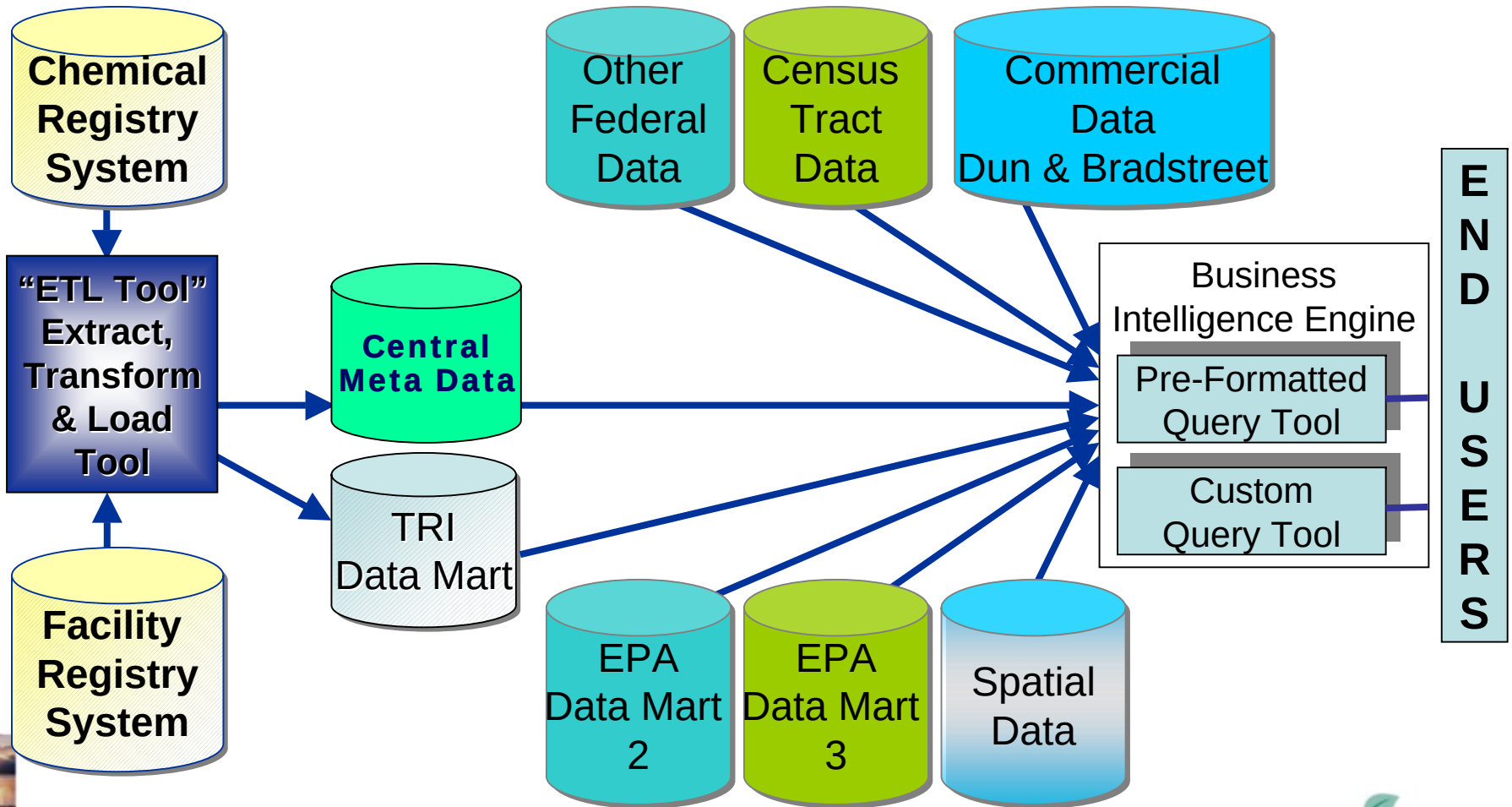
Star Schema
Facts and Dimensions

- Actual Data Marts
  - Air Quality + CAMD
  - TRI (Toxic Release Inventory)
  - Air Quest
- Actual Conformed Dimensions
  - Facility (FRS)
    - Monitor, Site (Lat/Long, Name, Street Addr.), Agency
  - Substance (Pollutant, Chem, Met.)
    - Protocol, Standard
  - Date/Time

- Planned Data Marts
  - Solid Waste
  - Watershed Initiative
  - Public Access
  - Enforcement
  - Indicators
  - Financial
- Planned Conformed Dimensions
  - Industry Sector
  - Demographics (Census)
  - . . .

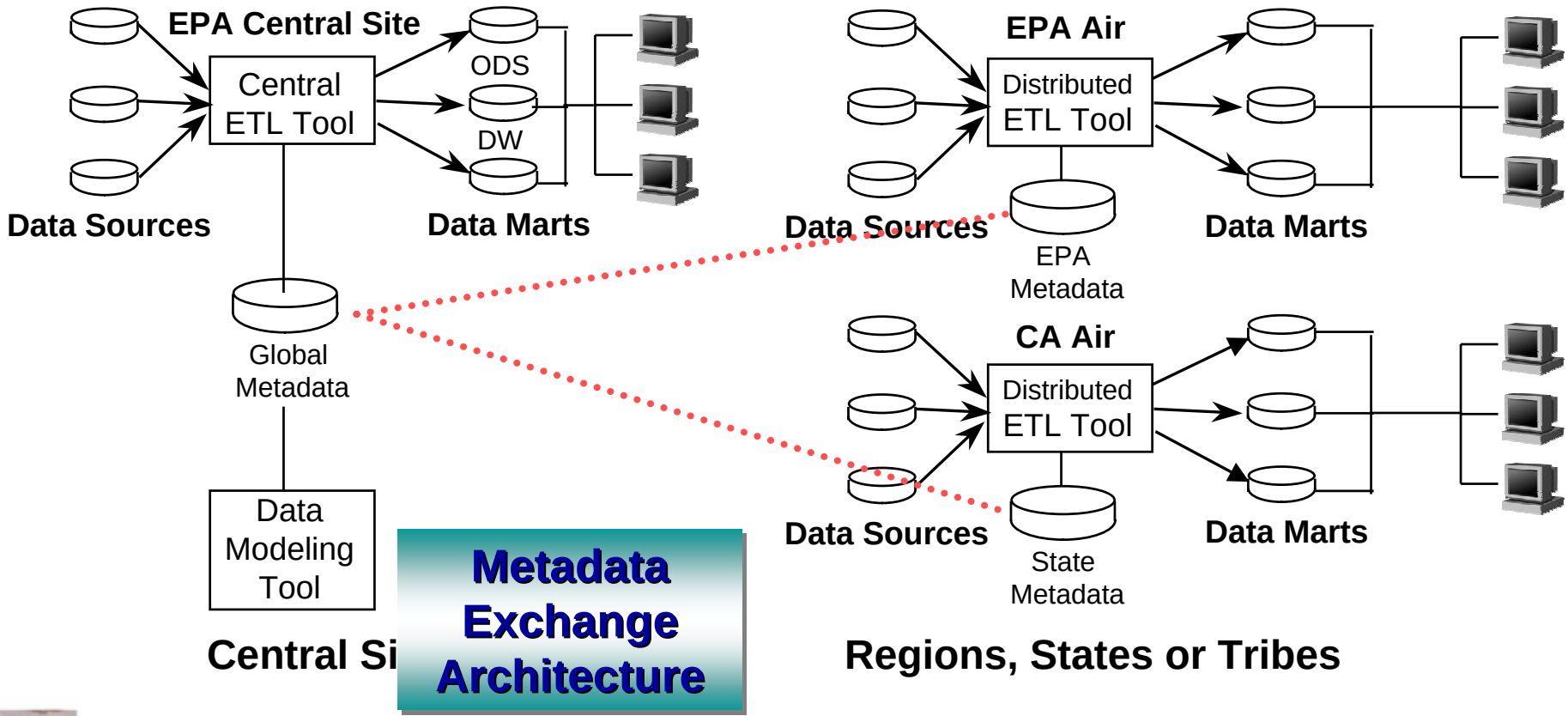
OFFICE OF ENVIRONMENTAL INFORMATION

Future Data Marts - Aggregation



OFFICE OF ENVIRONMENTAL INFORMATION

Future Data Marts - Franchising



Automation Goal: Eliminate 3GL Burden

- Reduce Exchange Network Partner Burden
  - Replace custom extracts/requests with ETL extracts
  - Implement CDX standards as business rules in ETL
- Reduce Envirofacts Burden
  - Replace custom extracts driving Envirofacts with ETL tool extracts
  - Reconcile multiple “standard” data exchange schemas
  - Leverage all available analyses of source data characteristics and use to specify ETL tool extracts
- Extend the Solution
  - Gradually extend the infrastructure to additional Domains
  - (Administrative Systems, Program Systems, and OEI)

ETL Tool Evaluation

OFFICE OF ENVIRONMENTAL INFORMATION

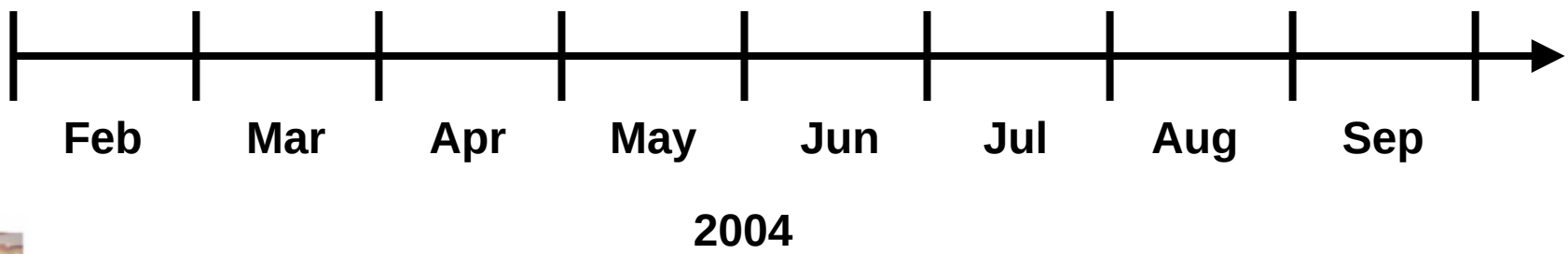
Objective

- To evaluate top-tier ETL tools (utilizing Agency-specific test cases in EPA's hardware and network environments) to determine their suitability for use as EPA's Data Integration Platform:
  - Technical capabilities
  - Company viability in the ETL tool market
  - Training and support services
  - Total Cost of Ownership

Critical Success Factors

- An AQS data mart and readiness to deliver future data marts
- Full synchronization with System of Registries and interface with OAR data warehouse
- Selection of ETL tool environment with standards and procedures for broad use within the Agency
- A pricing model documenting lower costs for subsequent data marts
- Explore and document linkage to business intelligence tools

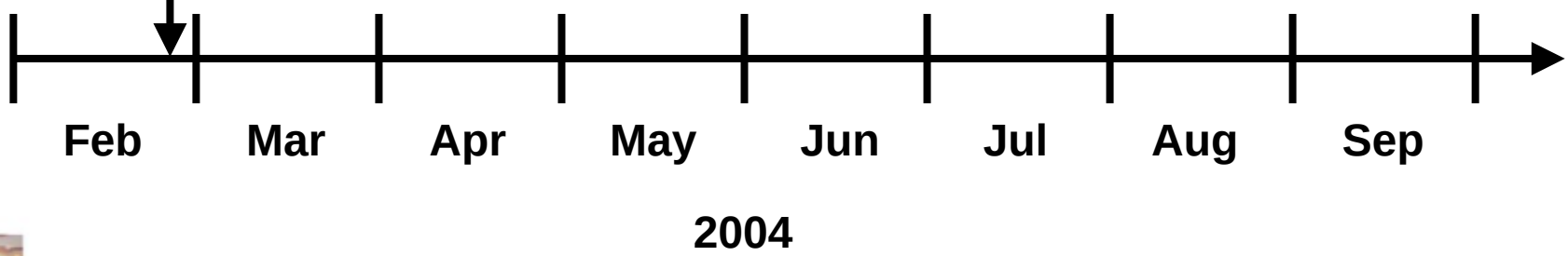
Methodology



OFFICE OF ENVIRONMENTAL INFORMATION

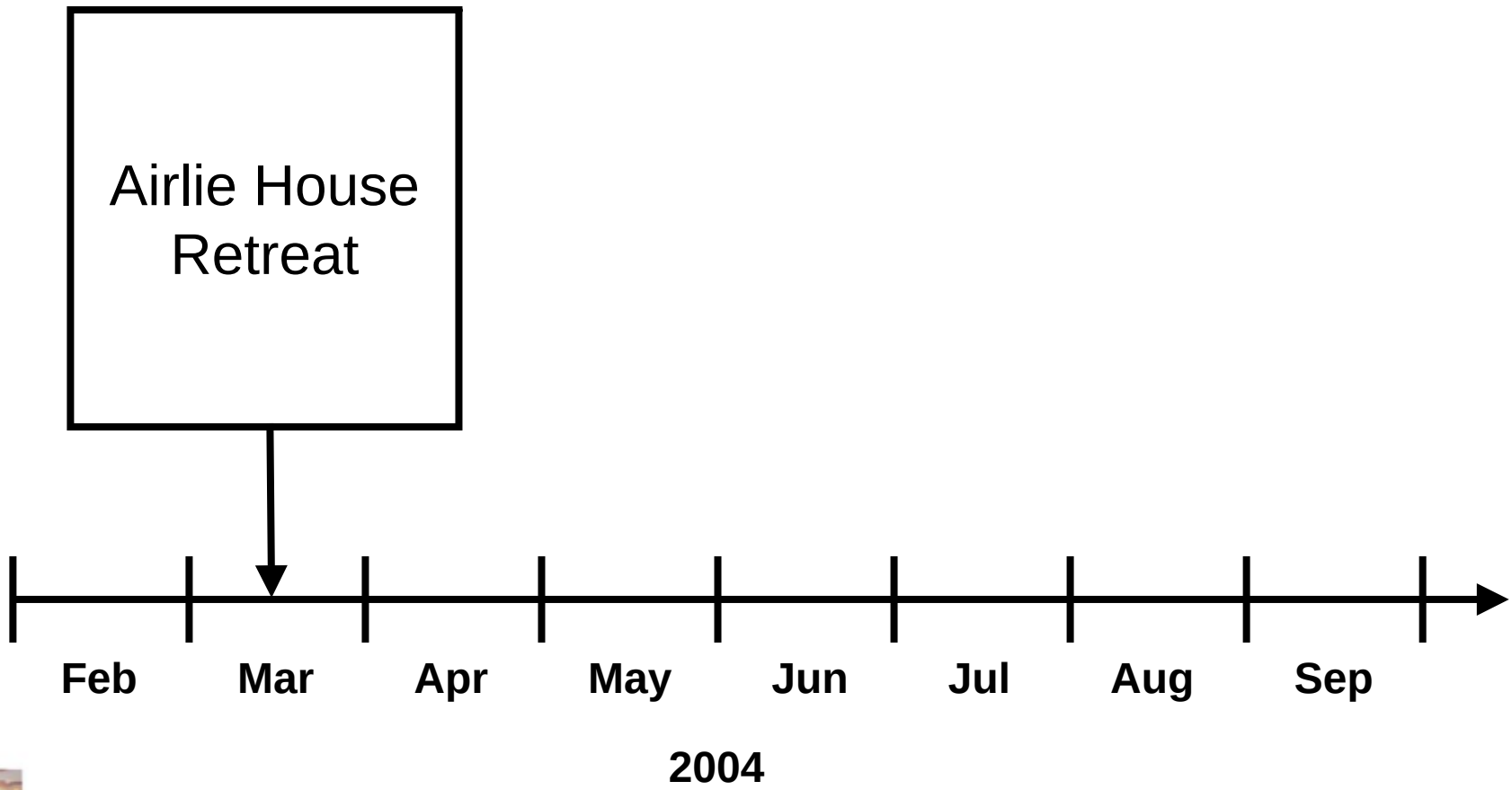
Methodology

PMO Program Office Forum on Data Integration Platform



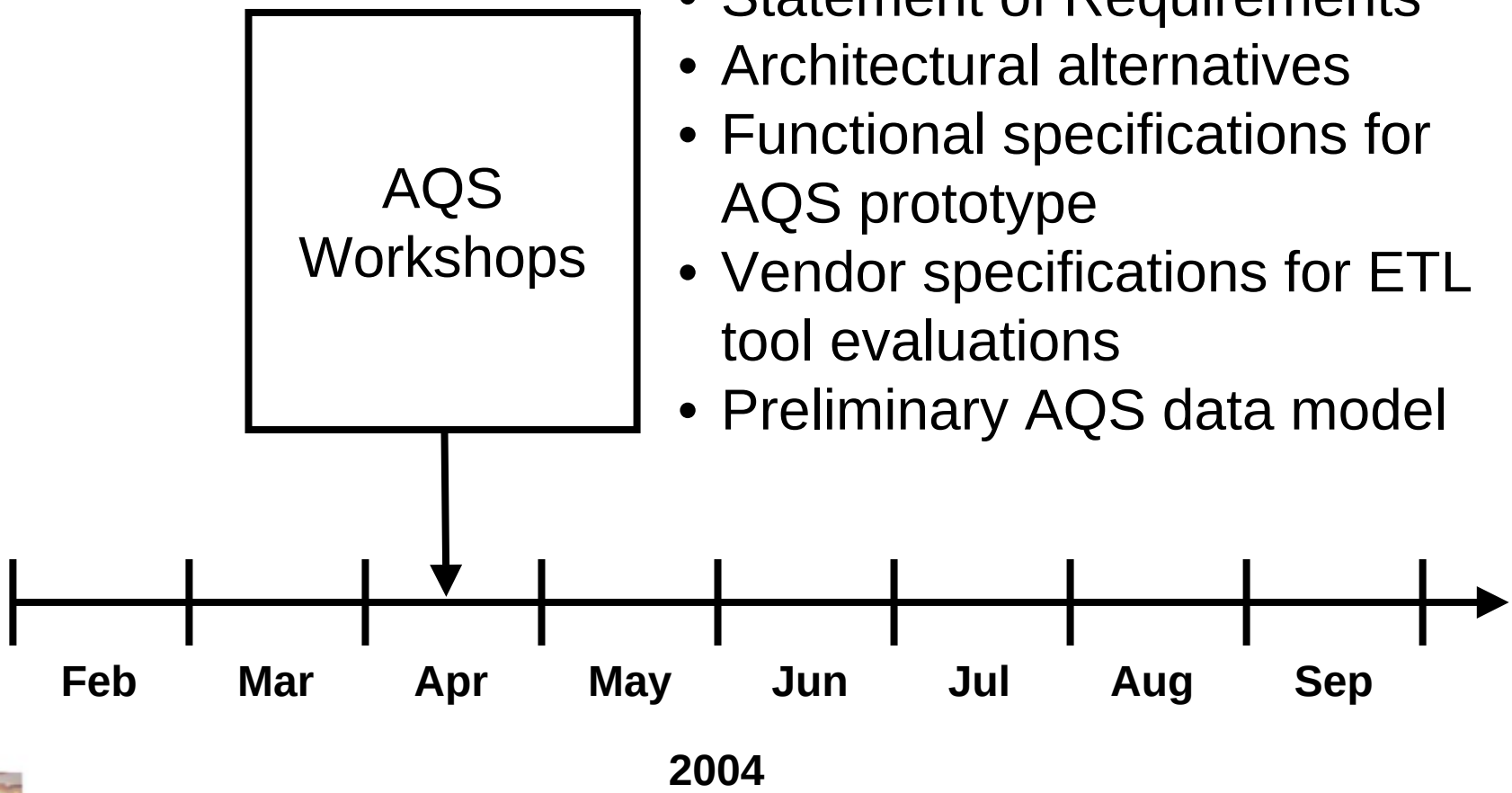
OFFICE OF ENVIRONMENTAL INFORMATION

Methodology



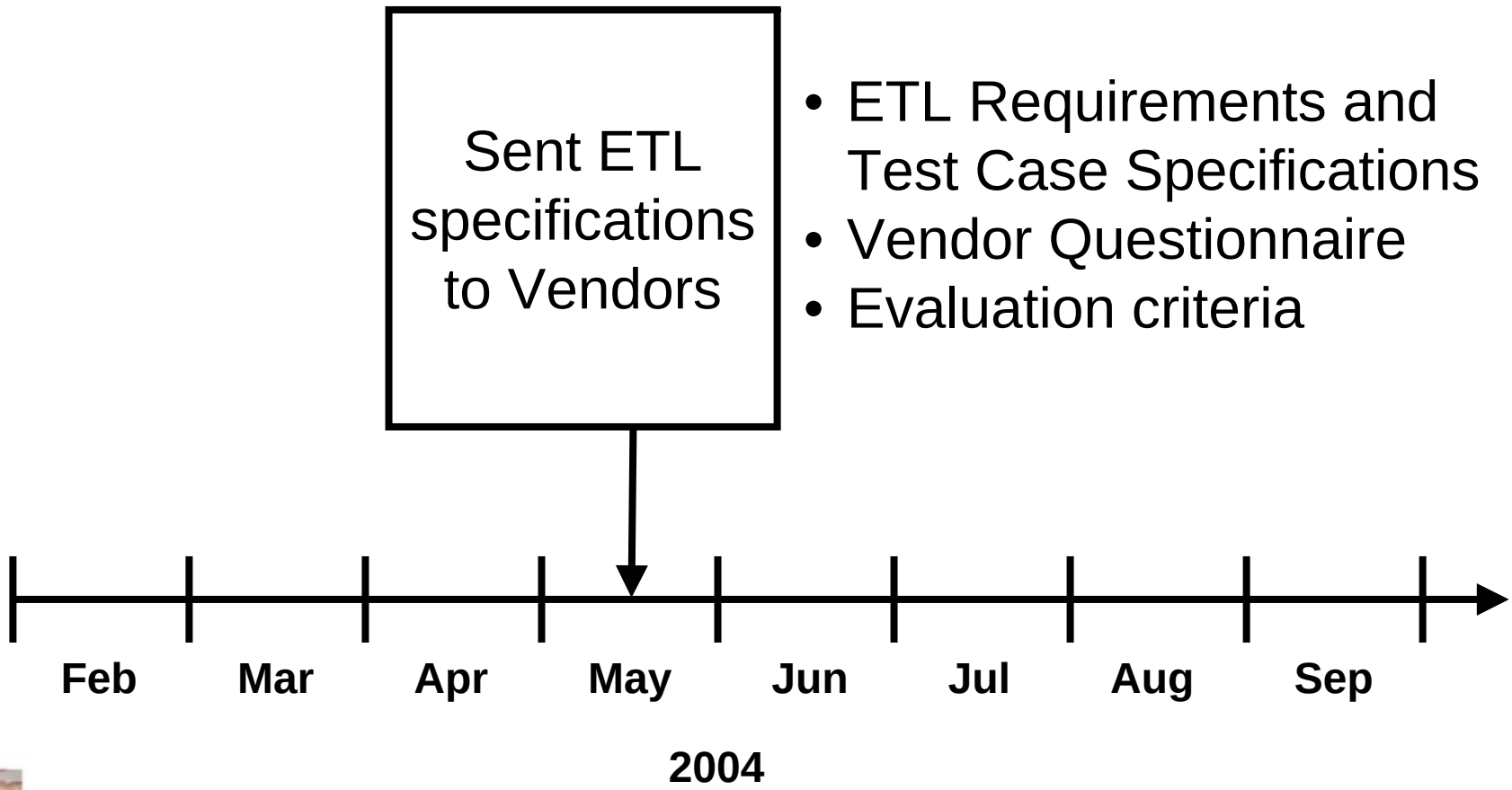
OFFICE OF ENVIRONMENTAL INFORMATION

Methodology



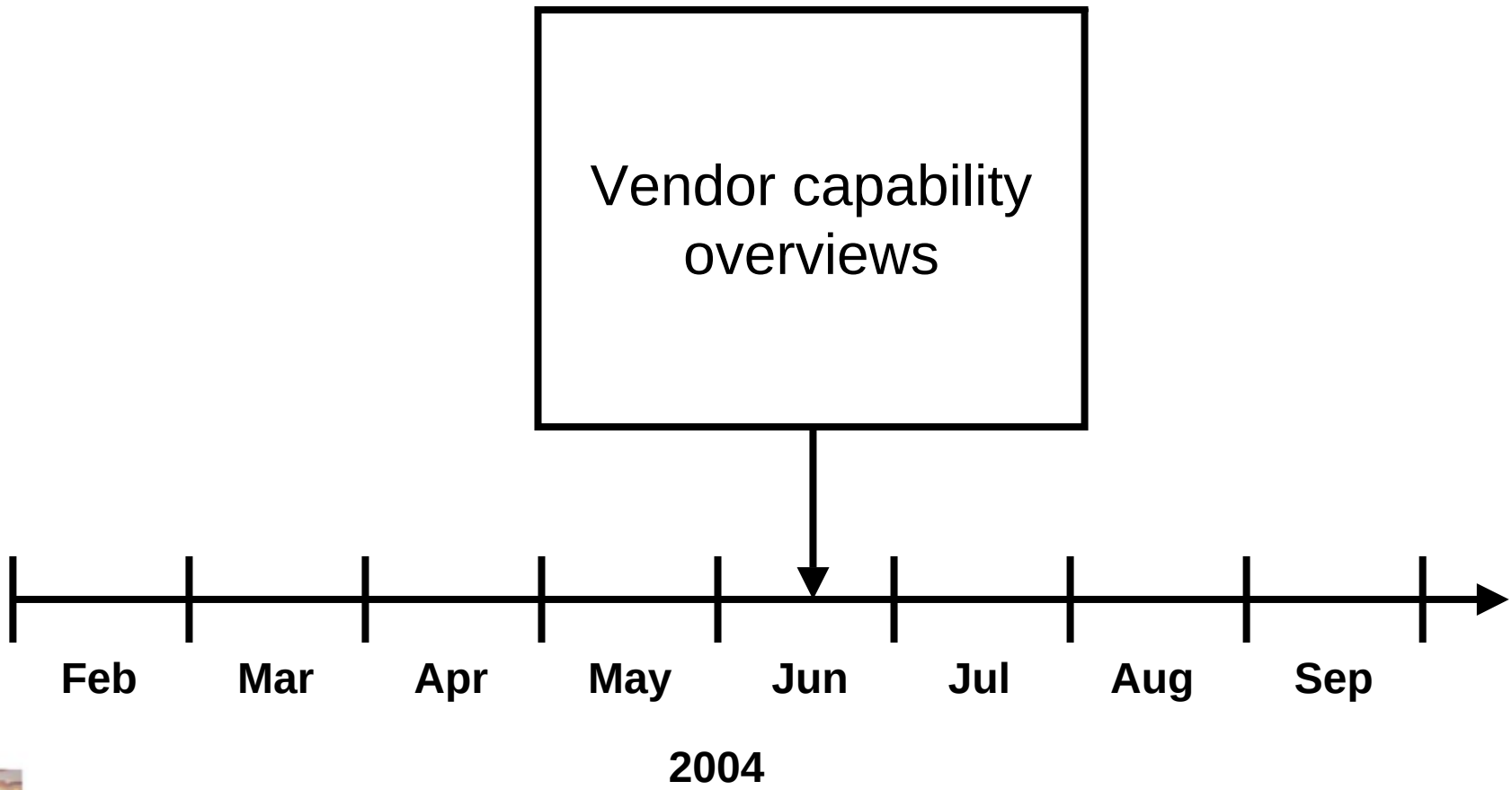
- Statement of Requirements
- Architectural alternatives
- Functional specifications for AQS prototype
- Vendor specifications for ETL tool evaluations
- Preliminary AQS data model

Methodology



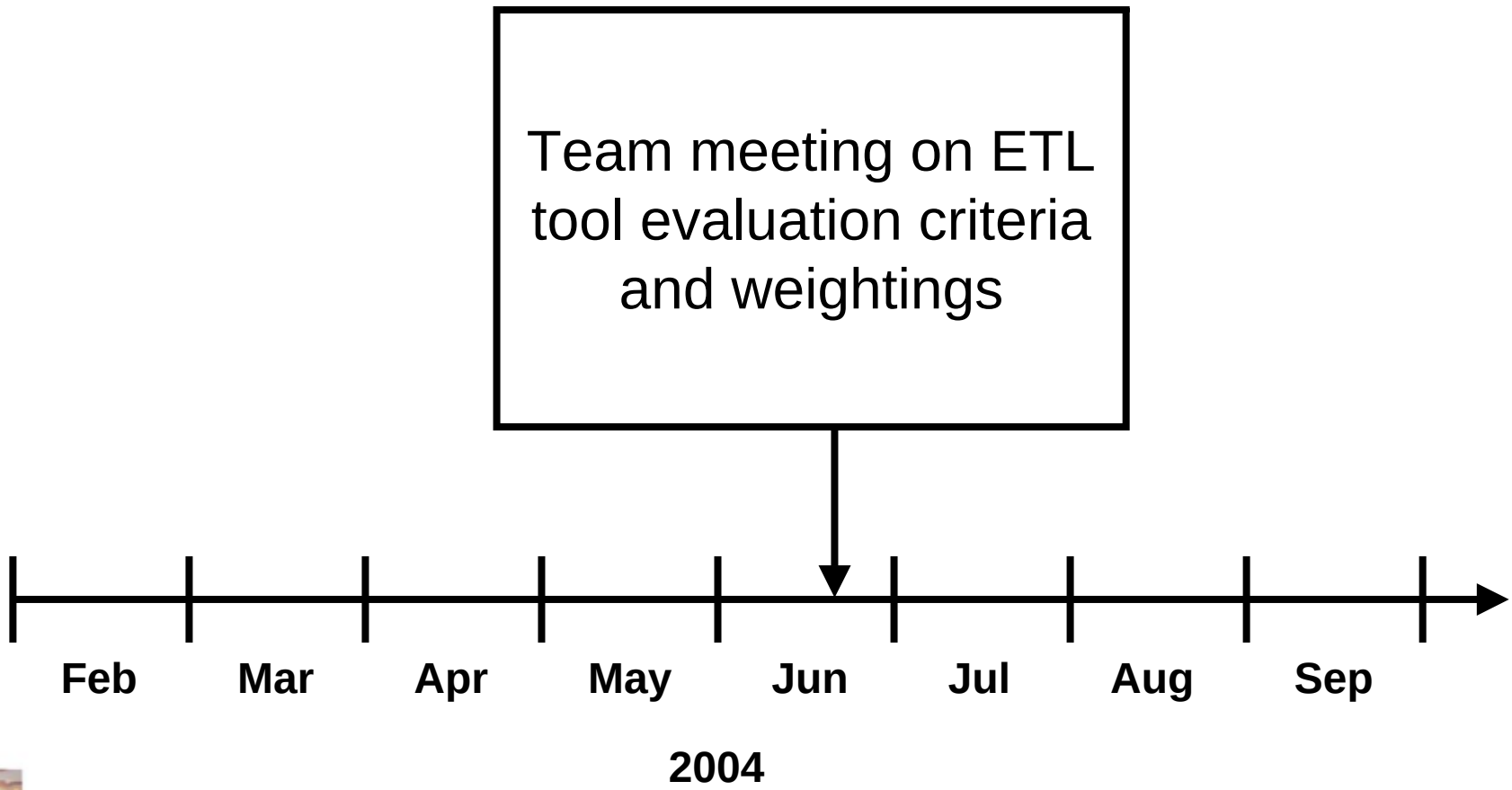
OFFICE OF ENVIRONMENTAL INFORMATION

Methodology



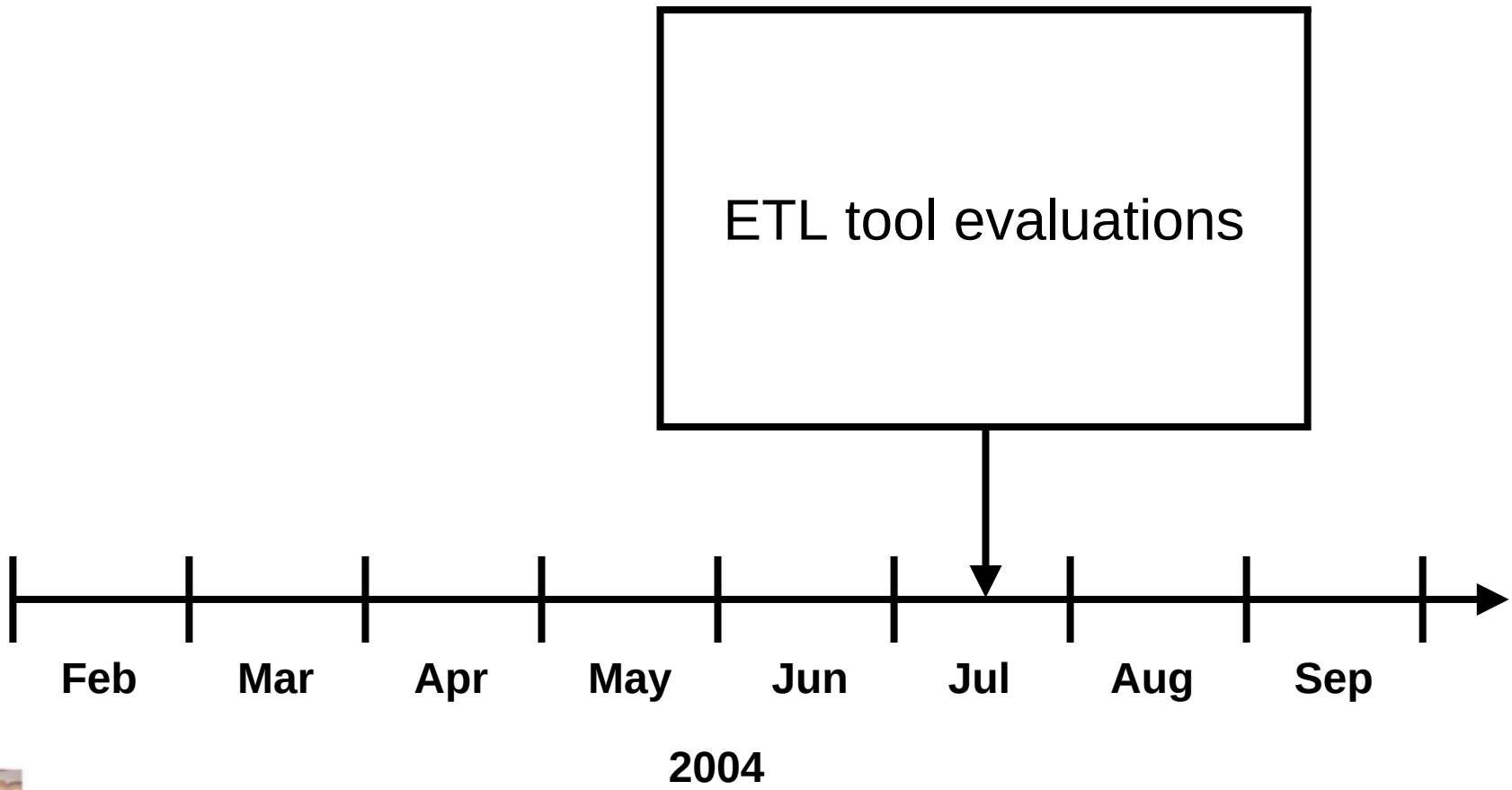
OFFICE OF ENVIRONMENTAL INFORMATION

Methodology



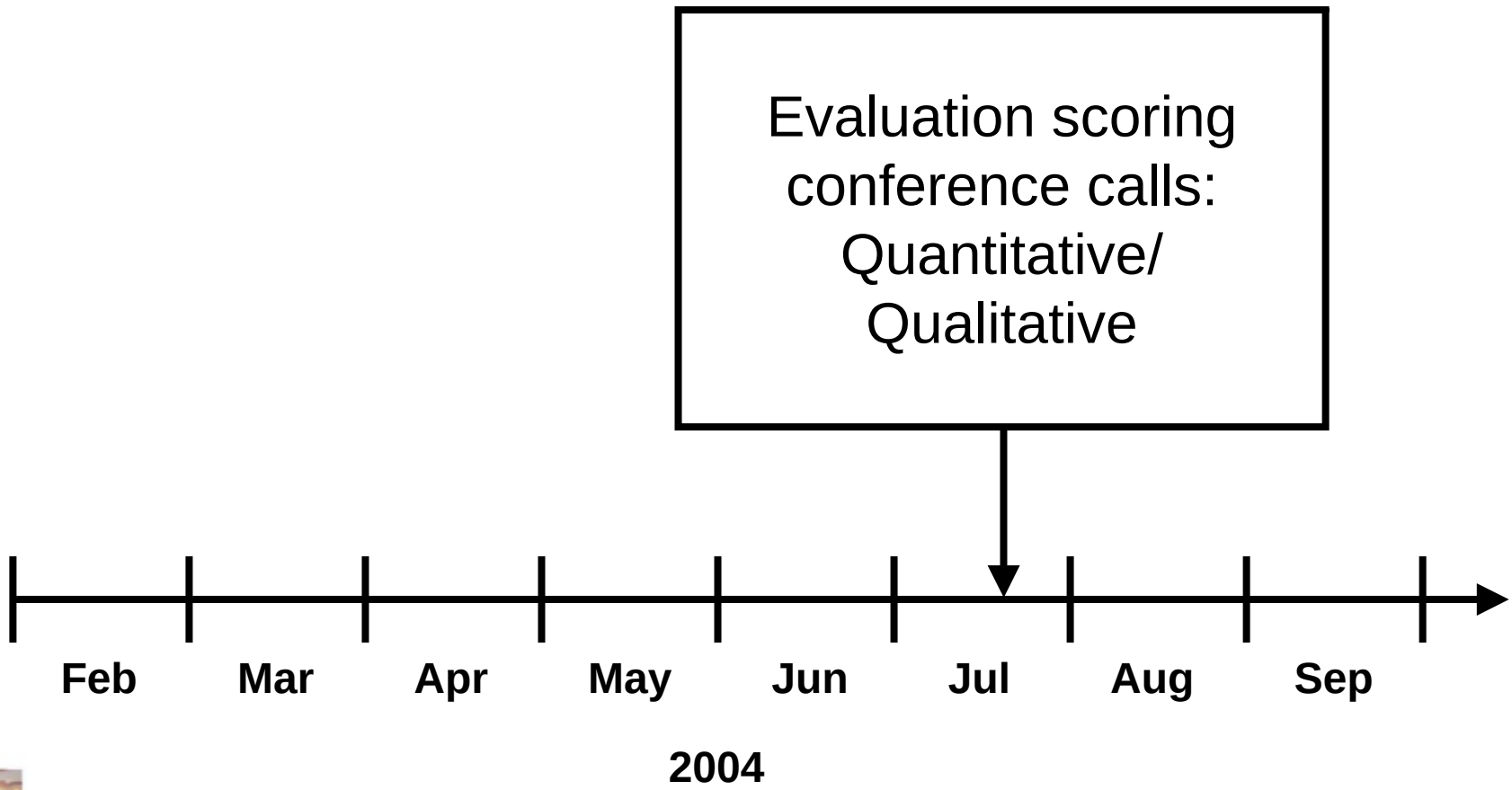
OFFICE OF ENVIRONMENTAL INFORMATION

Methodology



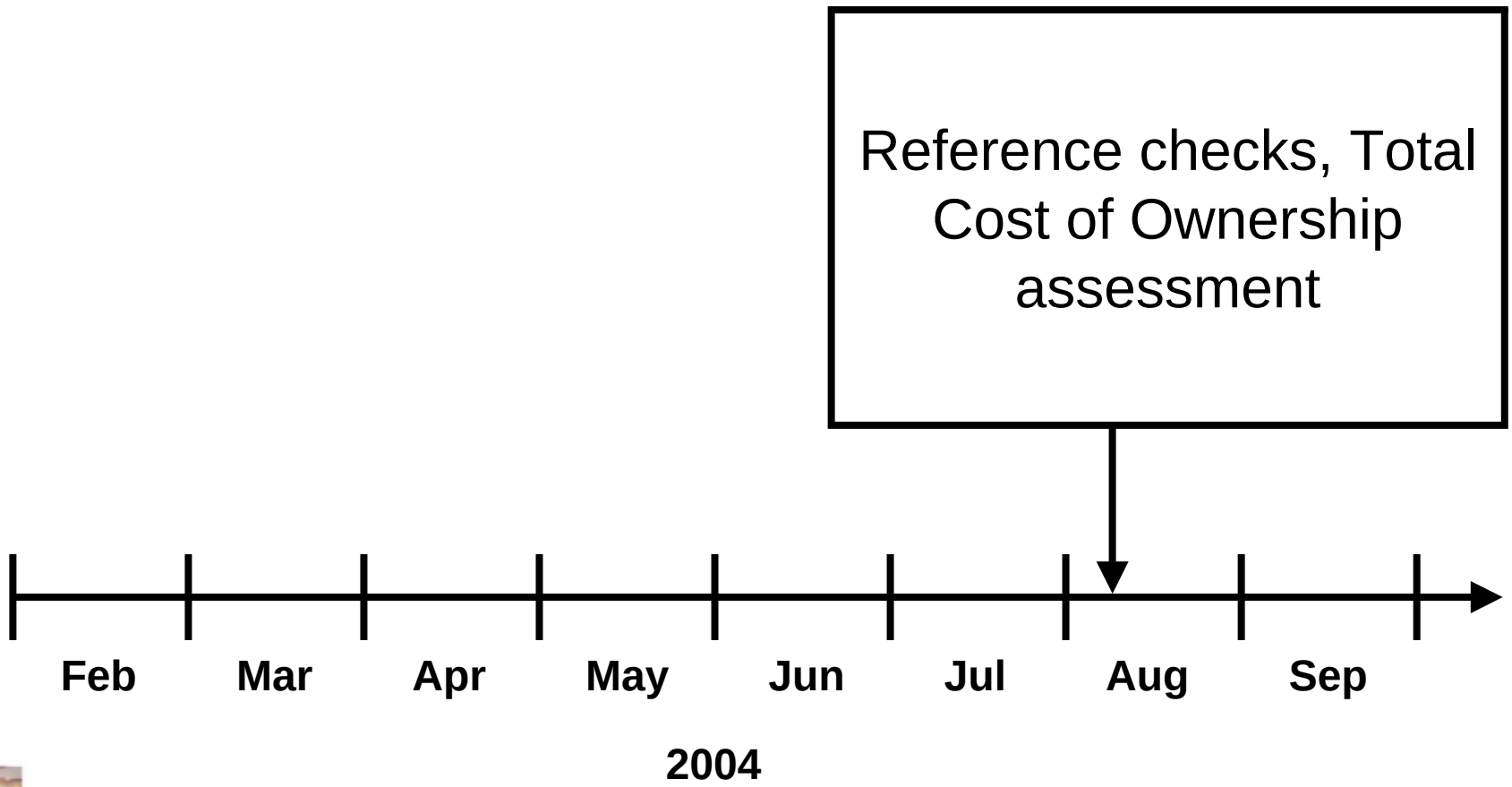
OFFICE OF ENVIRONMENTAL INFORMATION

Methodology



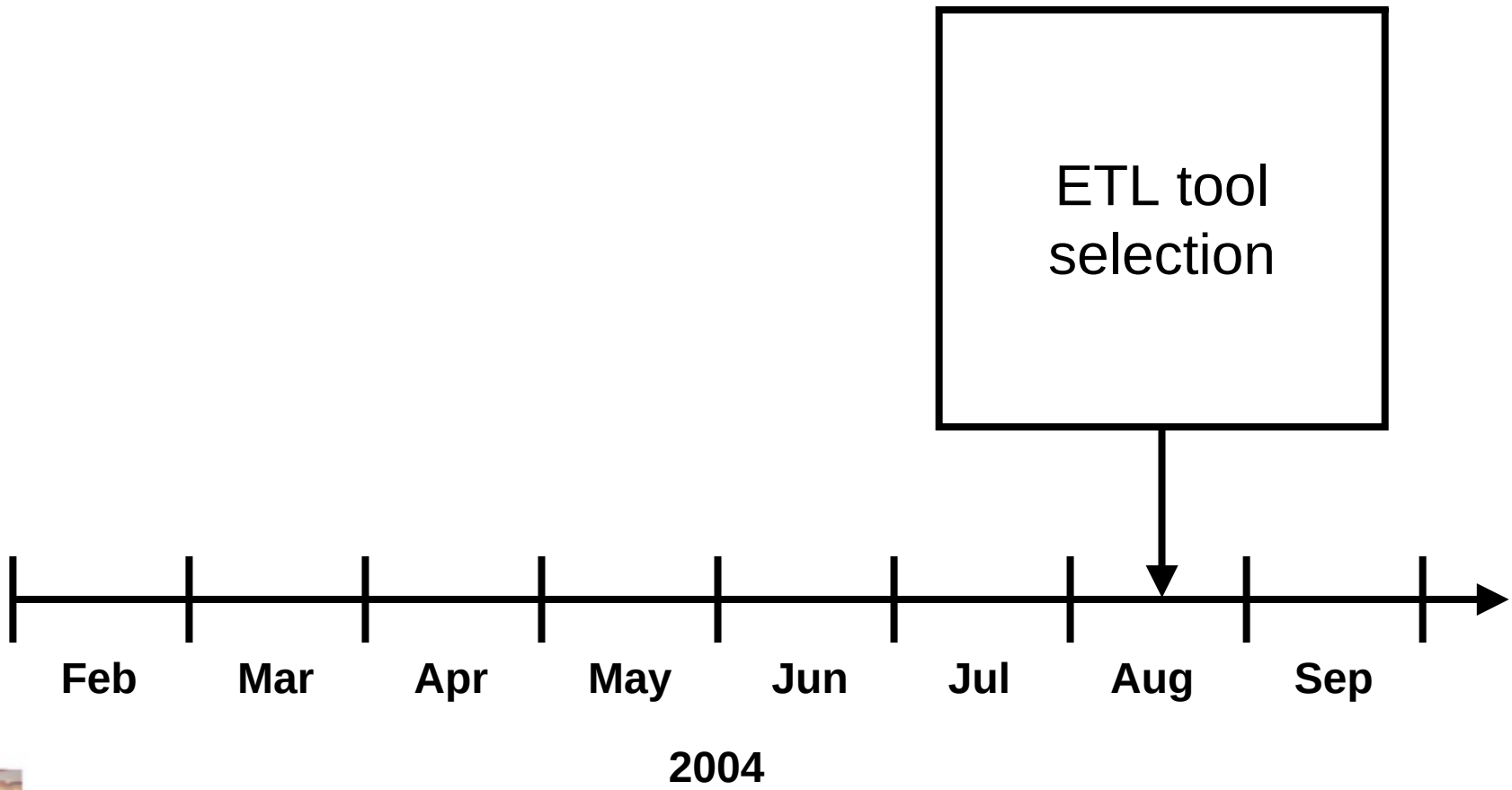
OFFICE OF ENVIRONMENTAL INFORMATION

Methodology



OFFICE OF ENVIRONMENTAL INFORMATION

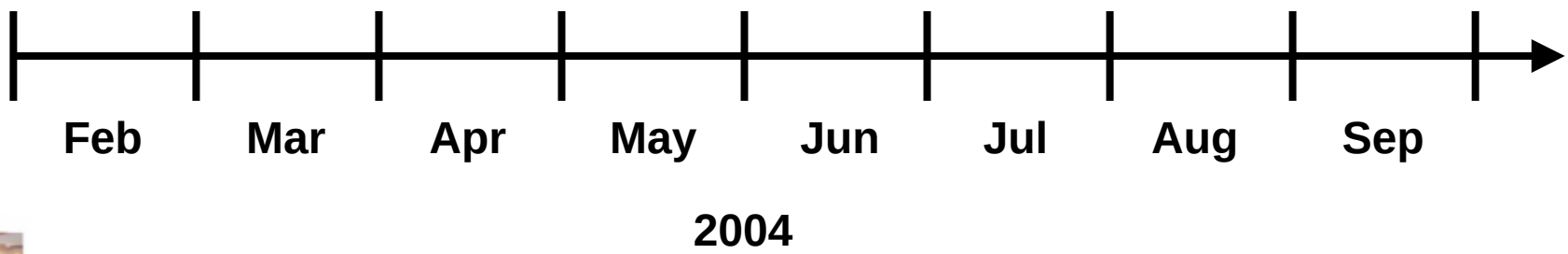
Methodology



OFFICE OF ENVIRONMENTAL INFORMATION

Methodology

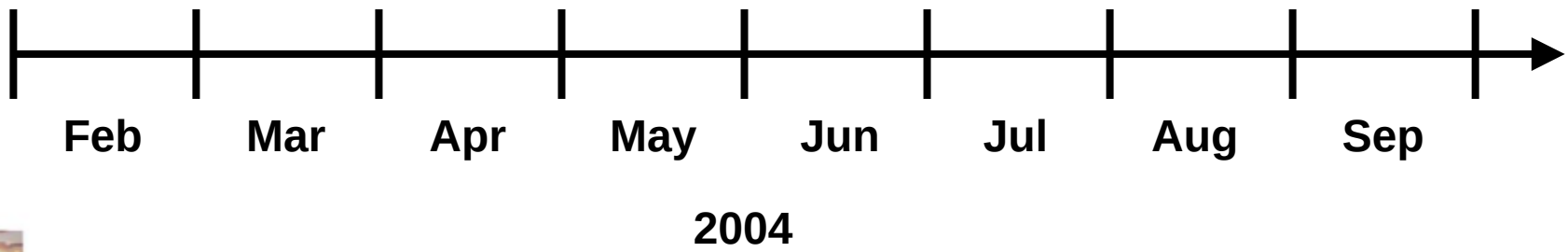
AQS data mart development – (90 day period following tool selection)



OFFICE OF ENVIRONMENTAL INFORMATION

Methodology

AQS production data mart:
(90 days following development)



OFFICE OF ENVIRONMENTAL INFORMATION

Total Cost of Ownership

- Areas of examination
  - Licensing and maintenance
  - Training/professional services
  - Hardware/infrastructure
  - Development costs

OFFICE OF ENVIRONMENTAL INFORMATION

Independent Process Review

- Independent source brought in to evaluate the methodology used in determining an ETL vendor for the EPA
  - Dr. Wang, Cambridge Research Group
    - Specializing in IT processes, methodology, and governance

Data Quality

OFFICE OF ENVIRONMENTAL INFORMATION

Data Cleansing Process

- Standardize EDR usage
  - Consistent “type” usage (Lat/Long, Substance)
  - De-duplicate facility addresses
- Clean Data
  - Load only “type” clean data
  - In the ETL process - if not at the source
- Measure Data Quality
  - Identify inconsistent, missing, incomplete, duplicative, or incorrect values
  - Build profiles to feedback to data stewards

Automated Data Profiling

***Data Profiling* is the assessment of data to understand its content, structure, quality and dependencies.**

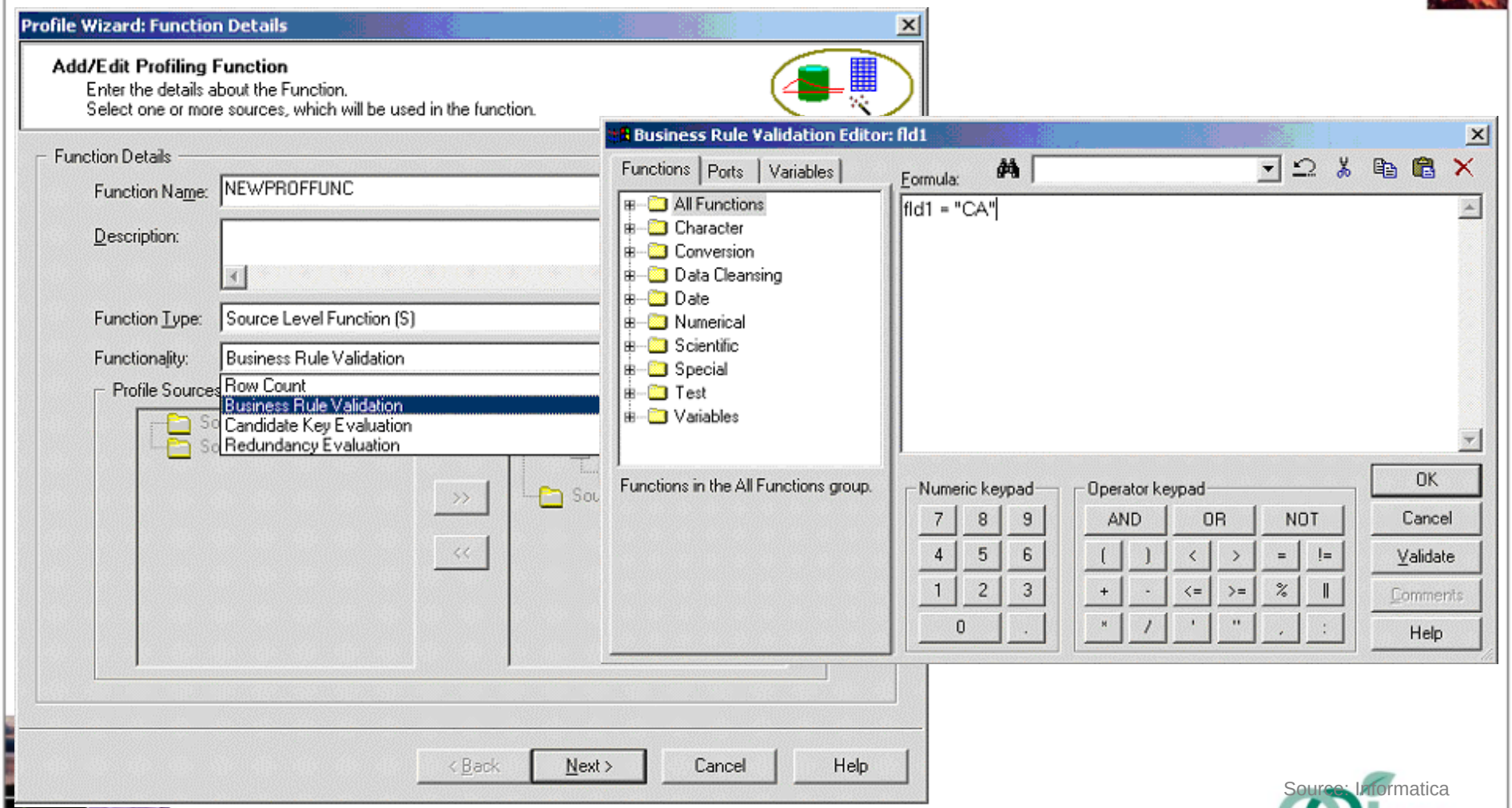
Column Level	▪ **Range validation, min/max, average, count** ▪ **Domain validation: e.g. pre-determined values**
Source Level	▪ **Row count** ▪ **Redundancy evaluation**
Intersource Level	▪ **Outer join analysis** ▪ **Cardinality analysis**

Automated Data Correction

***Data Correction* is the application of business rules and logic to fix data problems as an integral part of the data integration process**

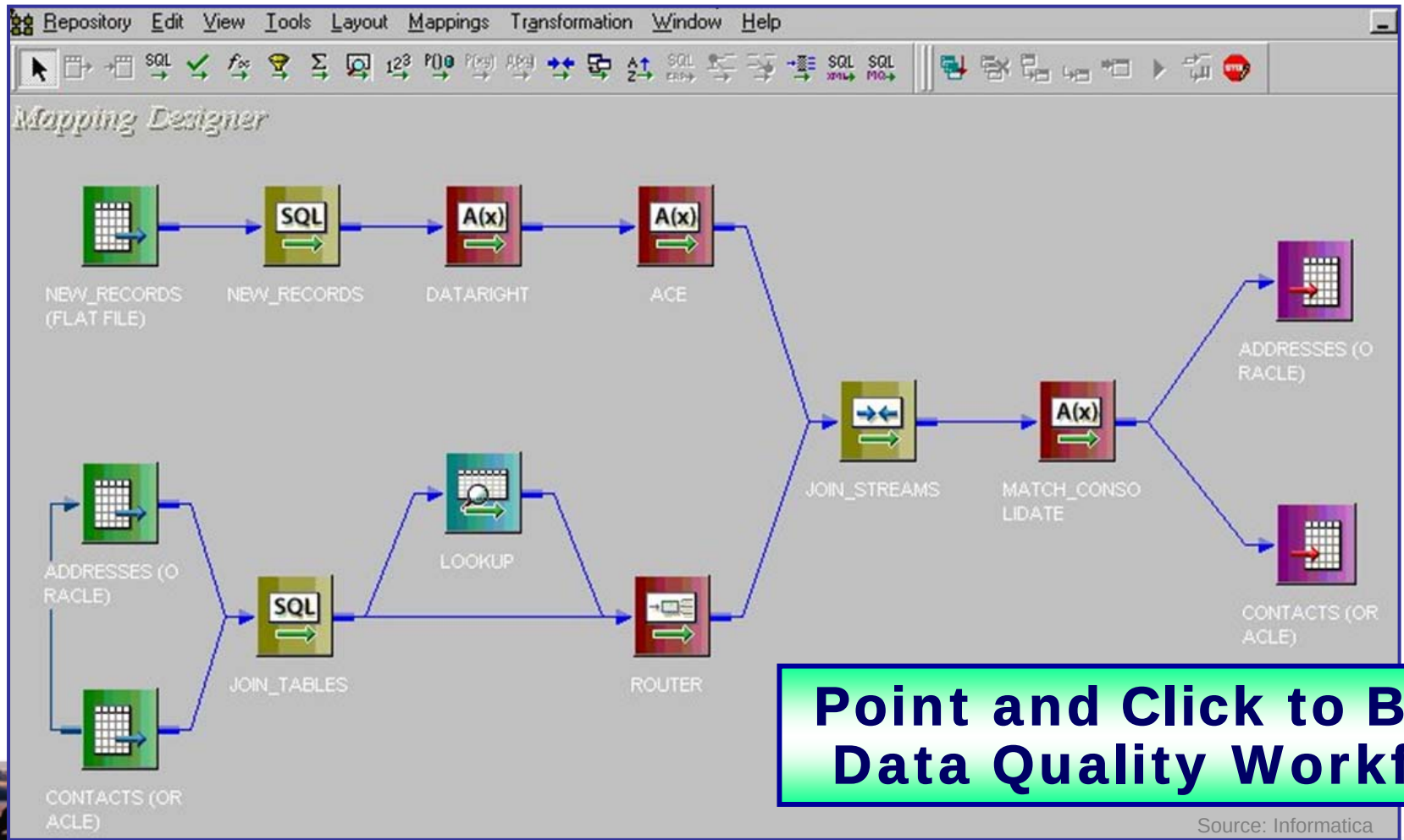
Transform	▪ **Convert data into different format, structure, and content to gain desired output**
Parse	▪ **Finds and breaks apart patterns within fields so parts can be processed independently**
Name & Address	▪ **Standardizes name and address data based on 3rd party data**

Automated Setup of Profiling or Correction Rule Data Quality Objects



Source: Informatica

Assemble Data Quality Objects into Automated Workflow



Point and Click to Build Data Quality Workflow

Source: Informatica

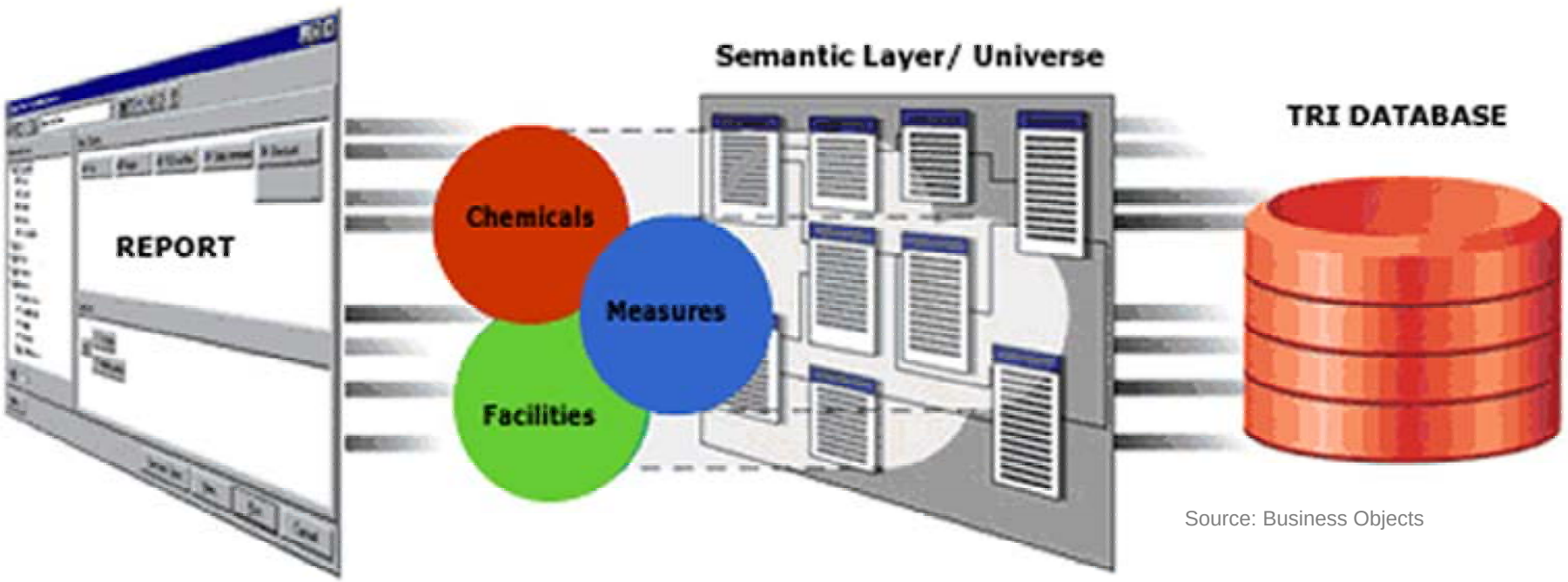
OFFICE OF ENVIRONMENTAL INFORMATION

Automating the User's View

OFFICE OF ENVIRONMENTAL INFORMATION

Data Mart – Concept of Operations

- **Shield End Users from Data Complexity**
  - **Semantic layer maps database complex metadata into business terms such as Chemicals, Totals/Measures, or Facilities**



Source: Business Objects

TRI InfoView Document Orientation

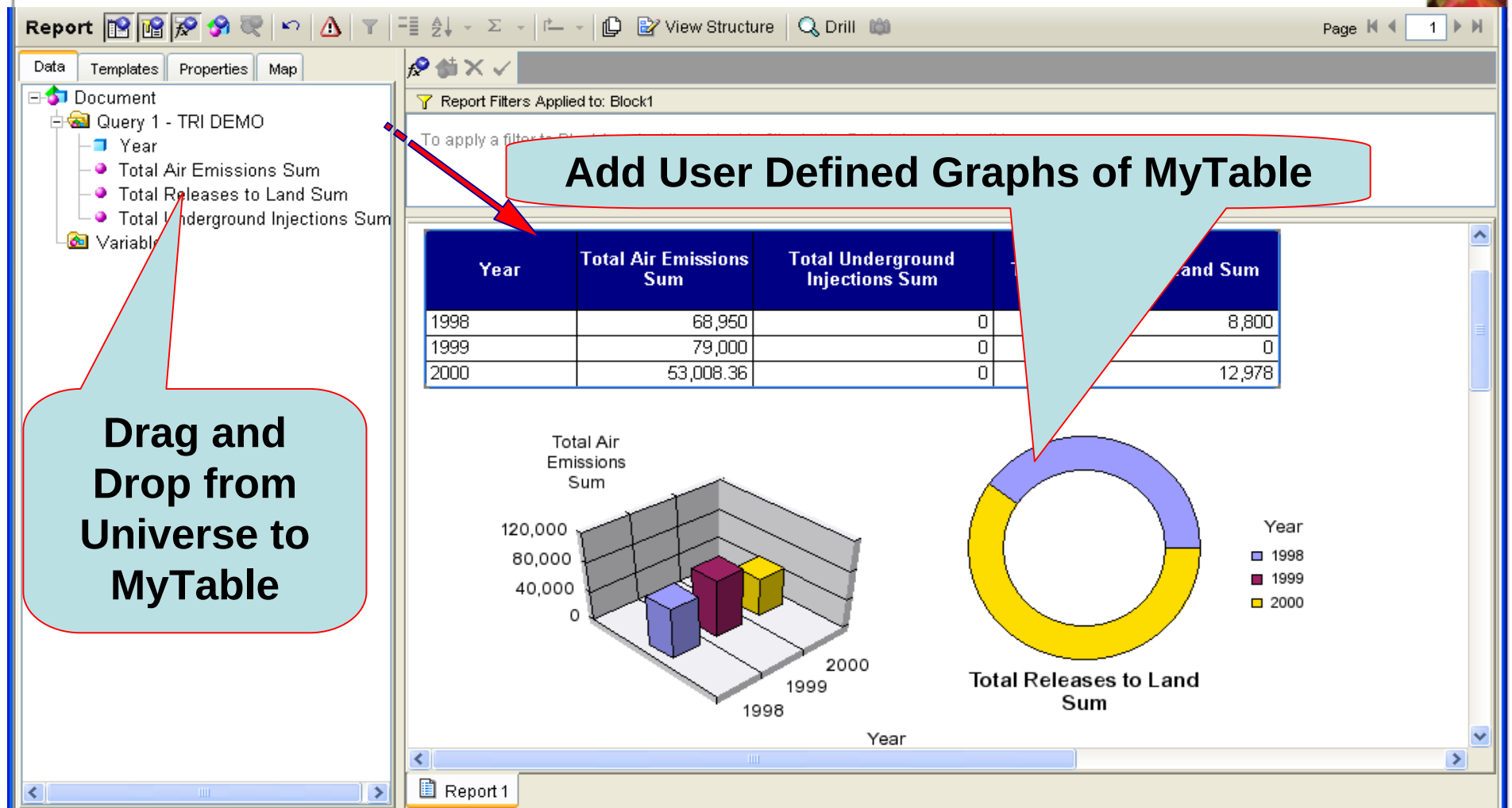
InfoView Query Results can be:

- **Saved on-line and shared among users.**
- **Scheduled for auto-generation and publication.**
- **Refreshed and saved as a new document for comparison.**

Example

OFFICE OF ENVIRONMENTAL INFORMATION

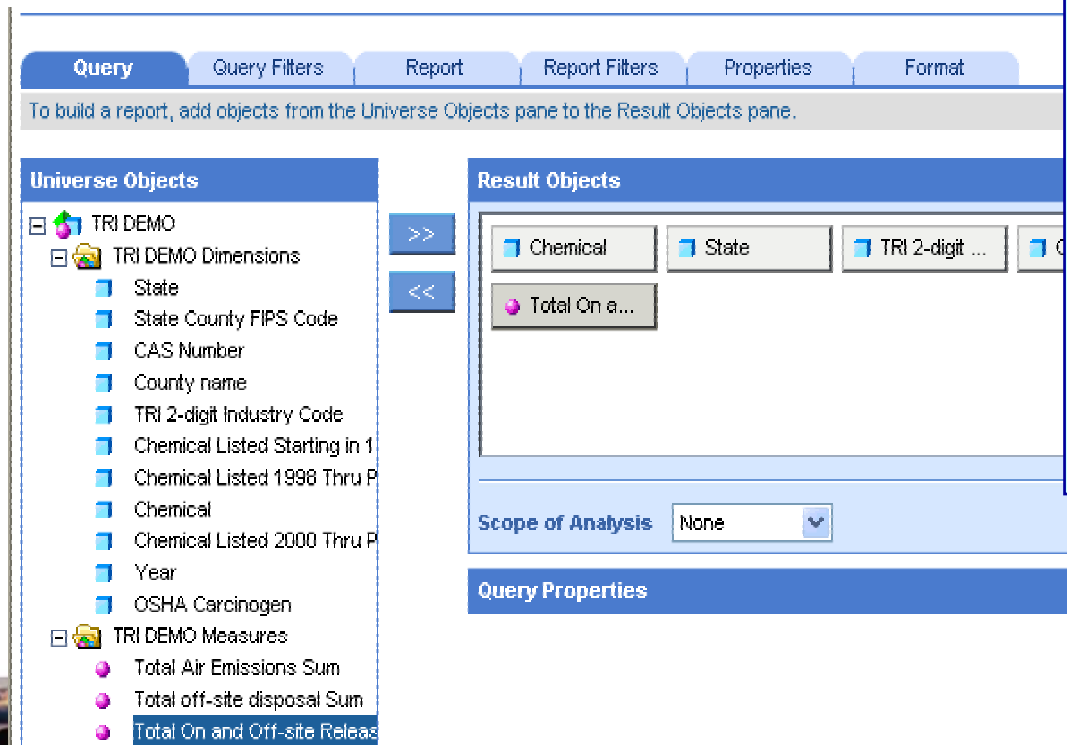
Build My_TRI_Report



OFFICE OF ENVIRONMENTAL INFORMATION

Customize My_TRI_Report

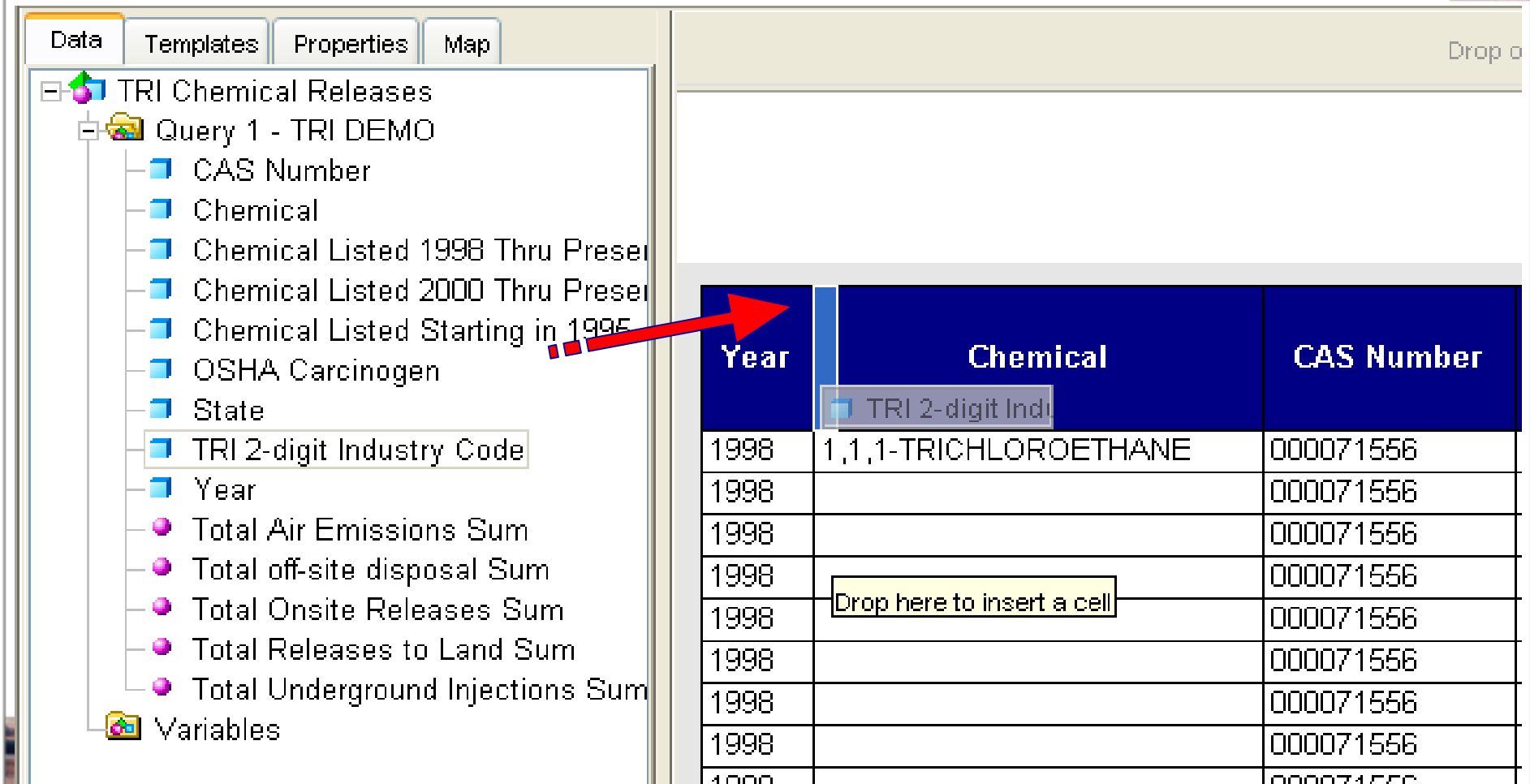
- Select Result Objects from Pallet of Business Objects and then…



- **Drag and Drop**
- **Group by**
- **Apply Filters**
- **Sorts**
- **Set Alerts**
- **Export**
- **Prompts**
- **Drill Down/ Drill Up**

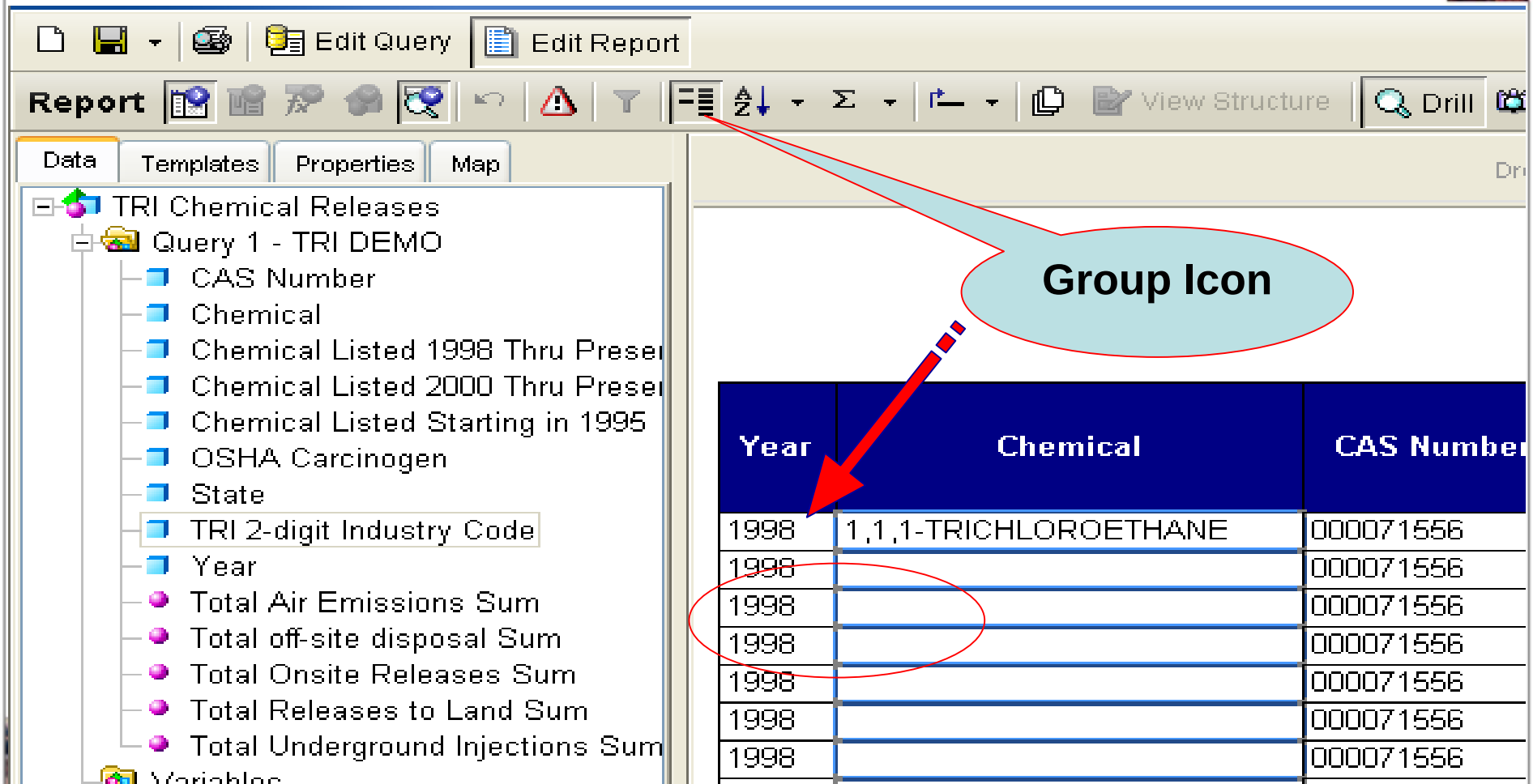
OFFICE OF ENVIRONMENTAL INFORMATION

Drag and Drop Attributes to Build a Report



OFFICE OF ENVIRONMENTAL INFORMATION

Group by Attribute



Apply Filters

TRI Toxics Release Inventory (TRI)

Edit Save Send View in PDF Format Add to My InfoView

Filter by

Document Map

TRI Chemical Release

- Report 1
  - 1,1,1-TRICHLOROETHAN
  - 1,1-DIMETHYL HYDRAZIN
  - 2,4-DINITROPHENOL
  - 2-ACETYLAMINOFLUORE
  - 4-AMINOAZOBENZENE
  - ACETAMIDE
  - AMITROLE
  - ANILINE
  - BENZENE
  - BIS(TRIBUTYLTIN) OXIDE
  - CARBARYL
  - CARBON TETRACHLORI
  - CHLORDANE

Year		nber	State	TRI 2-digit Industry Code	Che Li Star 1
1998	1,1,1-TRICHLOROETHANE	000071556	AL	73	
1998	1,1,1-TRICHLOROETHANE	000071556	AR	73	
1998	1,1,1-TRICHLOROETHANE	000071556	CA	73	

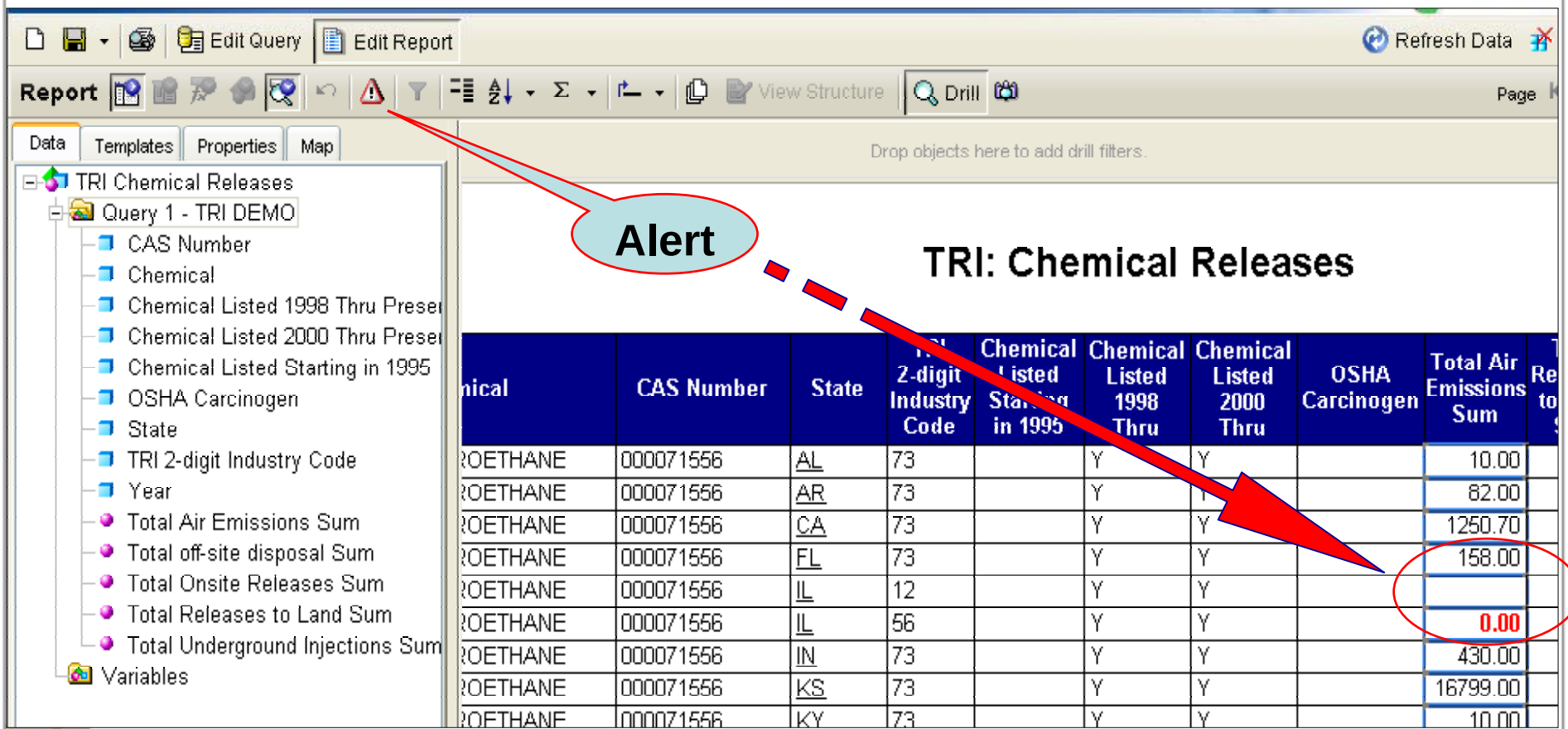
OFFICE OF ENVIRONMENTAL INFORMATION

Sort Function

Year	Chemical	CAS Number	State	TRI 2-digit Industry Code	Chemical Listed Starting in 1995	Chemical Listed 1998 Thru Present
1998	1,1,1-TRICHLOROETHANE	000071556	AL	73		Y
1998	1,1,1-TRICHL	071556	AR	73		Y
1998	1,1,1-TRICHL	071556	CA	73		Y
1998	1,1,1-TRICHL	071556	FL	73		Y
1998	1,1,1-TRICHL	071556	IL	12		Y
1998	1,1,1-TRICHL	071556	IL	56		Y
1998	1,1,1-TRICHL	071556	IN	73		Y
1998	1,1,1-TRICHL	071556	KS	73		Y
1998	1,1,1-TRICHL	071556	KY	73		Y
1998	1,1,1-TRICHL	071556	MA	73		Y
1998	1,1,1-TRICHL	071556	MI	73		Y
1998	1,1,1-TRICHL	071556	MN	56		Y
1998	1,1,1-TRICHL	071556	NE	73		Y
1998	1,1,1-TRICHL			73		Y
1998	1,1,1-TRICHL			73		Y
1998	1,1,1-TRICHLOROETHANE			73		Y
1998	1,1,1-TRICHLOROETHANE	000071556		73		Y

OFFICE OF ENVIRONMENTAL INFORMATION

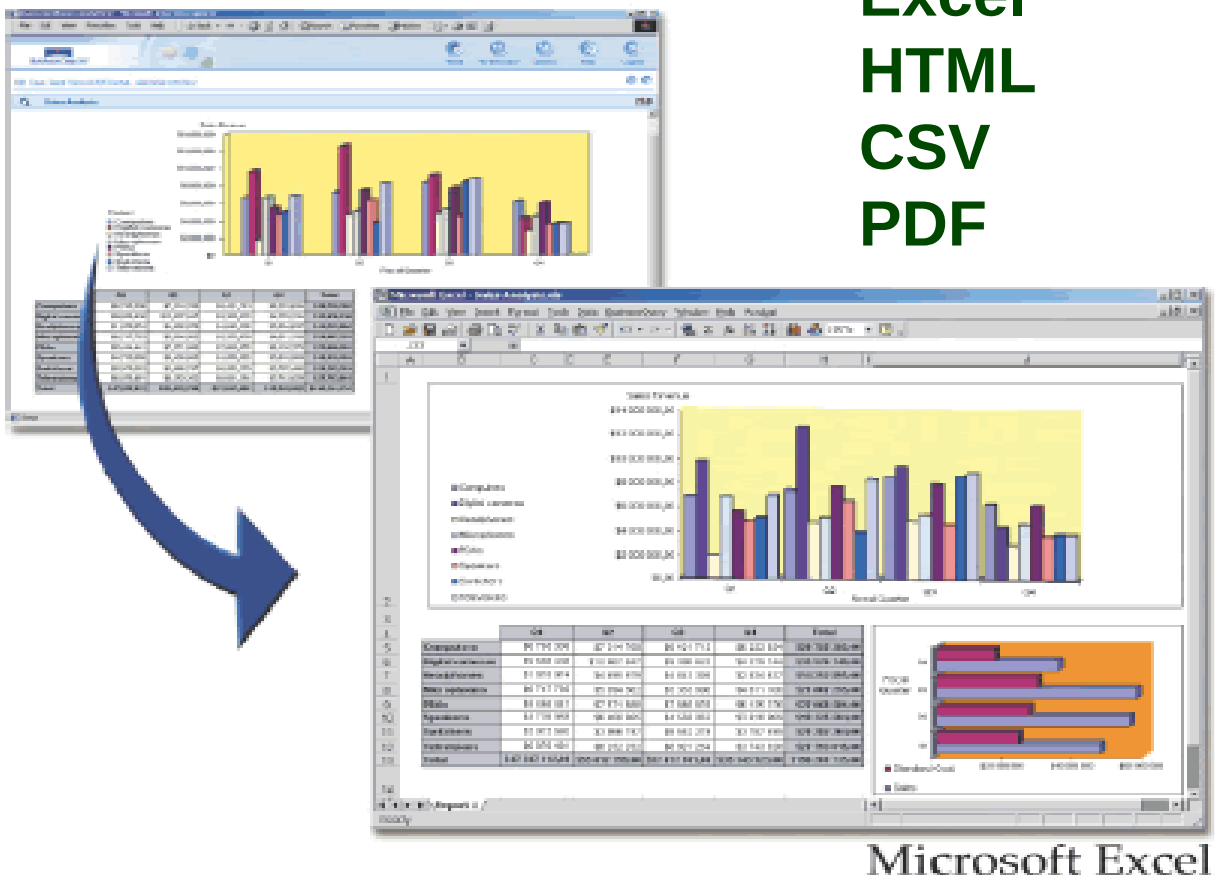
Alerts / Exception Highlighting



OFFICE OF ENVIRONMENTAL INFORMATION

Export to Favorite Format

- Export to any format



Source: Business Objects

OFFICE OF ENVIRONMENTAL INFORMATION

Summary

- EPA is Deploying the Next Generation of Data Mart Technology
  - Data Profiling, Data Cleaning, and Data Correction tools
  - Business Intelligence tools
- EPA is implementing the Next Generation of Data Bus Architecture
  - Star Schema / Reusable Conformed Dimensions
  - Distributed Data Aggregation
  - Franchise of Reusable Data Mart Methods and Data

Contact Information:

Bill Grabsch: grabsch.bill@epa.gov

OFFICE OF ENVIRONMENTAL INFORMATION