

Data and Information Management Research at the IRS

Victor J. Pérez-Núñez

703-983-7957 • vjperez@mitre.org

MITRE MSR

The logo for the MITRE Technology Program, featuring a stylized graphic of stacked blocks in yellow, orange, and blue to the left of the text.

**MITRE
Technology
Program**

Problem

- **The IRS has limited ability to:**
 - **Capture information and data for multiple inputs such as tax returns, and information returns**
 - **Analyze and provide timely reports from multiple data source systems**
 - **Promptly implement changes to business rules, caused by changing regulations, that could affect the processing of more than 40 terabytes of taxpayer data.**

Background: IRS Today

217 Million Returns

1.1 Billion Information Returns

80 Million Refund Checks

Returns, Payments Correspondence



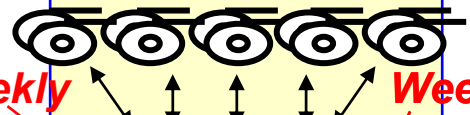
Capture/Perfect Data



W2s (SSA)
1099s



Master Files



Weekly

Weekly

Post/Settle Accounts

Daily

Active Accounts (10% Max)

~40 Terabytes

Financial Reporting (DCC)

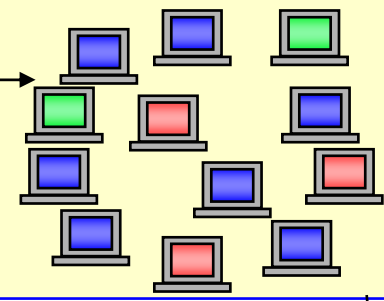
Revenue Accounting

IDRS



Integrated Data Retrieval

Functional Systems



Call Centers
Correspondence
Face-to-Face
Field Exam/Collection

Treasury, GAO

Weekly

Corporate Processing

Daily

Customer Service/ Compliance

Taxpayer (Telephone & Correspondence)

Main Characteristics: Complex Operations and Antiquated Systems

MITRE

Objective

- **Define a research agenda that will guide innovation grant studies dealing with information and data management (IM/DM) topics relevant to the IRS**
 - **Studies will help determine approaches to deal with the challenges to the IRS in managing the increasing information demand and volume of data for the current and future systems.**

Activities

- **Defined the following areas for evaluation:**
 - **Impact of XML to business processing**
 - **Enterprise metadata architecture and criteria to evaluate metadata tools and techniques**
 - **Impact of approaches to implement and manage business rules in an enterprise**
 - **Performance benchmarks for transaction processing and OLAP applicable to IRS environment**
 - **Application of benchmarks (e.g. TPC) given specific combinations of hardware and software algorithms for storage and retrieval**

Highlight

- **Defined research topics in the following areas:**
 - **XML: Study impact of XML compression methods on performance and storage (local and archival) requirements in e-filing environment**
 - **Metadata: Develop a bi-directional metadata architecture and define criteria to evaluate sets of integrated metadata tools and techniques**
 - **Business Rules Technology (BRT): Define approaches to develop and implement BRT in a large organization taking into account implications to the enterprise architecture**

Highlight/Demonstration

Receive XML Data

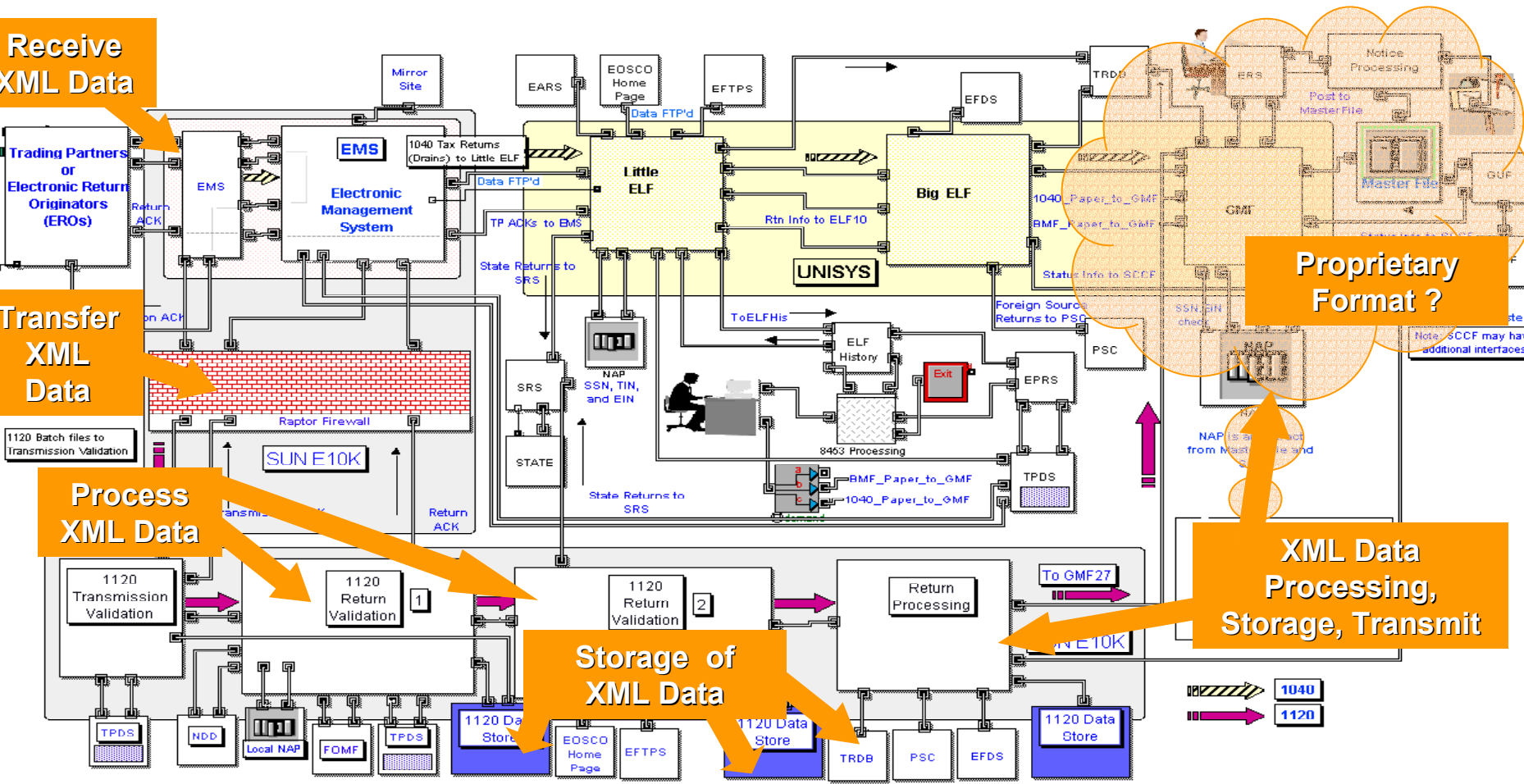
Transfer XML Data

Process XML Data

Storage of XML Data

Proprietary Format ?

XML Data Processing, Storage, Transmit



- Impact of XML compression approaches will be evaluated via modeling and simulation of the 1120 form processing

Impacts

- **As new components of the IRS modernization program are defined and developed in support of the goal to provide timely information to stakeholders, MITRE will be able to:**
 - **Directly relate and apply the research results of topics defined in this project**
 - **Influence the IRS adoption strategies of technologies such as XML, business rules, and metadata technology for information management**

Future Plans

- **Brief the IRS on specific topics in the research agenda to guide adoption of improved implementations**
- **Present research proposals to further develop identified topics addressing the IM/DM challenges of the IRS**
- **Develop an IM/DM framework that relates and facilitates alignment between the MITRE research and the IRS challenges in the work program**