Field Data Warehousing and Integration.
Big data, big problems, big solutions.
2017 ICT4D Conference

May 16, 2017

- **Catholic Relief Services**: Alvaro Cobo & Jeff Lundberg
- **Vera Solutions**: Caitlin Ferguson, Senior Consultant
- **Accenture**: Naveen Balani
Agenda – 90 minutes

• Panel Introductions (10 minutes)

• Accenture (20 minutes)
  • What is a data warehouse?
  • What does it look like?
  • What does the adoption of data warehousing in the non-profit sector look like?

• Vera Solutions (20 minutes)
  • Successful cases of data warehousing, integration and aggregation.
  • Best practices and (potential) standards for data integration.

• Catholic Relief Services (20 minutes)
  • What are the context and drivers for pursuing a solution?

• Q & A (15 minutes)

• Wrap-up (5 minutes)
What is a data warehouse?
- What does it look like?
- What adoption like in the non-profit sector?

May 16, 2017

- Accenture: Naveen Balani
What is a DATA WAREHOUSE?

A data warehouse is a subject-oriented, integrated, time-variant and non-volatile collection of data from one or more disparate sources.

Data warehouses store current and historical data and are used for analysis and delivering business intelligence.

**Key usages** –

- Business operations reporting
- Forecasting
- Multidimensional analysis
- Finding correlation among different factors
DATA WAREHOUSE - what do they look like?

Environments
A typical data warehouse environment comprises of
• ETL (Extraction, Transformation and loading solution)
• OLAP (Online Analytical Processing)
• Data mining software/solution
• Management tools

Design
• Design based on high level entities/subjects (i.e. customer, products)
• Schema based on facts and dimensions
Move towards MODERN DATA WAREHOUSE

**BASE INFRASTRUCTURE**
- Relational, Hadoop, Big Data, Streaming, Social, External

**QUERY ANALYSIS**
- ETL, Simple/Complex Query, Real-Time Stream Query, NLP

**BUSINESS INTELLIGENCE**
- Real-Time Reporting, Machine Learning, Predictive, Prescriptive

**Data Sources**
- Traditional
- Unstructured (Blogs, Images, Social..)
- IoT (Smart Devices, Cognitive)
Move towards MODERN DATA WAREHOUSE

**AMAZON REDSHIFT**
- its scalable columnar data store & tight integration with Amazon EC2 and S3 platforms

**MICROSOFT AZURE SQL DW**
- takes advantage of the Fast Track reference architecture and decouples of Storage & Compute

**GOOGLE BIGQUERY**
- is serverless delivering fast results without worrying about compute

**IBM DASHDB**
- distinguishes itself through its in-memory processing and in-data base analytics, uses Netezza analytics

**TERRADATA**
- provides its Data Warehousing, Aster discovery platform and Hadoop

**SNOWFLAKE**
- directly load structured & semi-structured data along with decoupling of Storage & Compute
What does the adoption of data warehousing in the non-profit sector look like?

<table>
<thead>
<tr>
<th>Areas</th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Availability</strong></td>
<td>Still a lot of data on paper or excels and not structured data</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Data Sharing across Non Profits</strong></td>
<td>Very low as Data is an important asset for receiving funding</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Integrated Data View</strong></td>
<td>Non existent. Data is in siloes.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Non Profit ICT Investments</strong></td>
<td>Low as ICT investment is viewed as a diversion from core purpose</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Digital Adoption</strong></td>
<td>Limited use of Digital technologies and infrastructures</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ICT Skills in Non Profits</strong></td>
<td>Low or non existent.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Data Governance</strong></td>
<td>No Standards</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Data Privacy and Security</strong></td>
<td>Low</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Decision Making</strong></td>
<td>Expert Judgment &amp; not Data Driven</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Transformation Journey</strong></td>
<td>Paper</td>
<td>Database</td>
<td>Data Warehouse</td>
</tr>
</tbody>
</table>
Field Data Warehousing and Integration. Successful Cases and Best Practices & Standards.

May 16, 2017

• Vera Solutions: Caitlin Ferguson
Data Solutions for Social Change

• Social enterprise with 190+ social impact clients in 45+ countries

• We believe...
  • In the power of data to make organizations more efficient & effective
  • That many organizations do not have the tools, systems, and/or capacity to do this well

• Focus:
  • Common challenges
  • Case study
  • Best practices
Common Challenges

• I have too many systems...

• I have a hard time reporting/aggregating my data...

• I don’t have access to the data that I need, or I need to rely on someone else for the data I need...

• I can only receive my reports on a monthly/quarterly basis...

• I’m not sure which data source is accurate ...
Case Study: PACT

• International NGO founded in 1971, operating in 30+ countries

• Challenges:
  • Business processes & data housed in silos
  • Multiple sources of truth
  • Data quality risks
  • Staff scrambled to assemble reports
  • Disparate financial systems
Case Study: PACT
Case Study: PACT

• Solutions:
  • Cloud-based, Salesforce system
  • Integration with NAV for key financial data
  • Uses Salesforce Apps
  • Manages 12,000+ project & contract docs

• Replaces manual work so staff can work on higher impact projects
• Improves knowledge management
• Collaboration by teams on same data
• Encourages curiosity, investigation, & analysis
Best Practices

• Identify goals
  • Which systems should be centralized, and why?
  • What is the value add for your organization?
  • What to integrate vs. what to eliminate?
  • What data to access, and who needs to access?

• Identify data points of entry and map data flows
• Define master system, and identify which platform best suits this
• Design for your needs; do not follow a template
Best Practices

• Data warehousing is not for every organization

• When is it not a good option?
  • If proposed systems for consolidation have no interaction with each other (e.g., data and/or Users)
  • If there’s no value gained from consolidating (e.g., if staff’s workload will be more challenging or increase)
Best Practices

• If data warehousing is for your organization:
  • Plan, plan, plan!
  • Carefully consider integrations, data migration, & data deduplication
  • Look at the whole picture: it’s fine to build in phases, but be sure to have your road map

“Most importantly [the system] provides a single version of truth in an organization that in the past had at least a dozen [systems].”

- Mark Reilley, Senior Director, Global IT, PACT
Field Data Warehousing and Integration. Project context and drivers.

May 16, 2017

- Catholic Relief Services: Alvaro Cobo & Jeff Lundberg
What do we mean by Field Data Warehousing?

Field Data Warehousing is...

• Storing our data
• Accessing our data
• Structuring our data
• Extracting our data
What is driving CRS to look at Field Data Warehousing?

- Open Data Policy
- Program Performance Metrics Initiative
- Identified need by CRS Global MEAL Community
- MEAL / Supply Chain Data Initiative
- Implementation of the MEAL Policies
- Digital Datasets from multiple tools
- ICT4D Strategy Discussions
- University linkages

Data Management
Business Case – Putting it all together!

Summary of Business Case – three compelling reasons to act:

1. Project-level M&E data is stored at the project level and is generally unavailable to those outside the project.
   - CRS’ knowledge is stored and proliferated across many platforms – or stored on local hard drives
   - Difficult to locate or lost/deleted upon program closeout – limiting learning opportunities agency-wide
2. There are increasing demands (both internal and external) on CRS to aggregate and share structured data

- Databases tend to be created each time a new project is initiated
- CRS’ programmatic M&E data is fragmented across many systems, collected in different ways, and structured in different formats.
- USAID’s open data policy and CRS’ Program Performance Measurement initiative require improved data management systems and processes that rise above the project level.
Business Case – Putting it all together!

3. Poor data management systems and practices expose CRS, its partners, and its beneficiaries to risk
   - Donors are increasingly demanding adherence to international data privacy and security standards
   - The lack of standardized data management systems and processes limit the agency’s ability to mitigate this risk
Putting it all together – What are the project goals and objectives?

**Project Goals**
- Improve accessibility of field project M&E data for data-driven decision making & program improvement.
- Improve quality of future grant proposals.
- Begin to provide the ability for the agency to comply with current & upcoming open data policies.

**Project Objectives**
- Provide a user friendly & scalable mechanism for selected Agriculture pilot projects to store & load agriculture program performance data into a data warehouse.
- Provide data visualization / reporting for selected Agriculture pilot projects.
- Improve agency capability to run agriculture program performance analysis on M&E data sets for selected Agriculture pilot projects.
Enable Monitoring & Evaluation (M&E) data to be an organizational asset to improve program quality for beneficiaries CRS serves by being deliberate of how M&E data is stored.

SMART Objectives

- Determine requirements and implement data co-location (i.e. a place for projects to put their raw data) to store M&E data by end of September 2017.
- Determine requirements for data warehouse by end of December 2017.

FY16 Q4 - Present

- August 2016 - Present
  - Strategic Planning
  - Vision Planning
  - Obtain Business Lead (Product Owner) for project
  - Define data standards for Agriculture pilot sector

FY17 Q4

- July - September
  - Determine requirements for data co-location (i.e. a place for projects to put their raw data)
  - Execute requirements for data co-location
  - Begin determining requirements for data warehouse

FY18 Q1

- October - December
  - Finish determining requirements for data warehouse
  - Choose solution based on requirements

FY18 Q2 to Q4

- January - September
  - Plan & implement solution
  - Begin pilot of solution in selected Agriculture projects

Fun Fact:

- Data co-location is different than a data warehouse.
- The former is in scope for implementation in 2017.
Audience & Panel
Q&A