

**A Distributed Collaborative Workflow Based Approach To Data  
Collection and Analysis**

William Gerecke, Douglas Enas  
Raytheon Company  
6225 Brandon Avenue, Suite 230  
Springfield, VA 22150  
[gerecke@rayva.org](mailto:gerecke@rayva.org), [denas@rayva.org](mailto:denas@rayva.org)

Susan Gottschlich  
Raytheon Company  
528 Boston Post Road  
Sudbury, MA 01776  
[sng@rayva.org](mailto:sng@rayva.org)

# **A Distributed Collaborative Workflow Based Approach To Data Collection and Analysis**

William Gerecke, Douglas Enas  
Raytheon Company  
6225 Brandon Avenue, Suite 230  
Springfield, VA 22150  
[gerecke@rayva.org](mailto:gerecke@rayva.org), [denas@rayva.org](mailto:denas@rayva.org)

Susan Gottschlich  
Raytheon Company  
528 Boston Post Road  
Sudbury, MA 01776  
[sng@rayva.org](mailto:sng@rayva.org)

## **Abstract**

Data Collection and Analysis (also known as After Action Review) capabilities are common requirements for many Command and Control (C2) and Modeling and Simulation (M&S) systems and architectures. In our work we have found that in order to be maximally effective, these capabilities must be designed with the military user workflow process in mind. In this paper we present a web-enabled Data Collection and Analysis capability that 1) considers a typical military workflow whereby several users, of varying levels of technical sophistication and disparate responsibilities, will need to make use of these capabilities, 2) addresses the need to enable distributed collaboration and 3) is based on a modular multi-layered service oriented architecture so that the same distributed collaborative workflow based approach can be used to satisfy a wide range of Data Collection and Analysis needs and can enable machine-to-machine interaction by exposing web services.

## **1. Introduction**

Data Collection and Analysis (DCA) is used in many different situations. In M&S applications, DCA systems are used to monitor and present Measures of Performance (MOP's) and Measures of Effectiveness (MOE's) of ongoing war games, course of action analyses, etc. In tactical training exercises, DCA systems are used to evaluate the performance of trainees and provide report cards to commanding officers. DCA is needed to support After Action Review (AAR) activities following military exercises.

Most recently, we have used our DCA system to monitor the health and performance of the C2 enterprise architectures that Raytheon built for U.S. Central Command (CENTCOM) and the Coalition Provisional Authority (CPA). An earlier version of our DCA capability was installed on the CENTCOM Deployable Headquarters (CDHQ) networks at Camp As Salayah, Qatar and used to analyze, monitor and troubleshoot problems with the CDHQ enterprise networks prior to the Internal Look '03 exercise and prior to Operation Iraqi Freedom. Our DCA systems have also aided analysis initiatives associated with the CPA enterprise networks we built for Baghdad.

Performing DCA for C2 enterprise architectures has given us a unique opportunity to experiment with a Data Collection domain that is well-defined and easy to implement thereby giving us an ideal harness to evaluate DCA systems as part of a typical

headquarters-level military workflow. This evaluation is further supported by the mission of a military headquarters – several coalition partners, directorates, and components come together at a headquarters and the DCA system must be able to support all of their needs.

From our experience in working with CENTCOM and the CPA, we have found that the network-centric split-based operation nature of modern warfare and peacekeeping place new requirements on DCA systems:

- Data Collection and Analysis is needed for all phases of an operation or exercise: for troubleshooting systems and architectures during the build up period, for performing health monitoring and MOE/MOP monitoring during the actual operation or exercise, for maintaining, tabulating, and evaluating historical trends for extended operations, and for developing and supporting After Action Reviews.
- DCA capabilities are needed by a wide range of users who have varying levels of understanding of the systems or processes being analyzed and of computational analysis techniques. For instance, a network router subject matter expert (SME) and a Joint Operation Center (JOC) chief will have disparate expertise and needs.
- DCA capabilities need to be accessible anywhere at anytime by any person using any device. While there may be restrictions placed on what data an individual user is allowed to access at a given time, data and analysis access should not be limited by the capabilities of the system or device. Distributed analysis capabilities are needed to support distributed collaboration.
- DCA extends beyond the traditional data collection domains needed to support MOE/MOP evaluation during an exercise. By developing a modular service oriented architecture, disparate categories of data can be collected by specially designed well-encapsulated data collectors, known as Data Adapters, so that the basic DCA system can be used to accommodate a wide range of uses and users.
- DCA users must be able to create new analysis products on the fly without code modification. For instance, if a new MOE or status chart is needed, a user must be able to create a template for the chart and publish it on the site for others to refresh (repopulate with the latest data) and view.
- DCA cannot be a standalone system. It needs to expose its data, results, and analysis capabilities, via web services, to other C2 systems.

## **2. DCA Architecture**

The modular, multi-layered, service oriented DCA architecture we have developed is depicted in Figure 1. It is based on a Microsoft's ASP.NET technology and makes heavy reuse of COTS and Open Systems standards, software, and code samples. The goal is to focus development efforts on domain-specific requirements and to utilize ASP.NET features to provide infrastructure such as the database management server, the web server, and the application server. An added benefit to this architecture is that technology refresh is facilitated because the architecture is necessarily modularized along industry established boundaries.

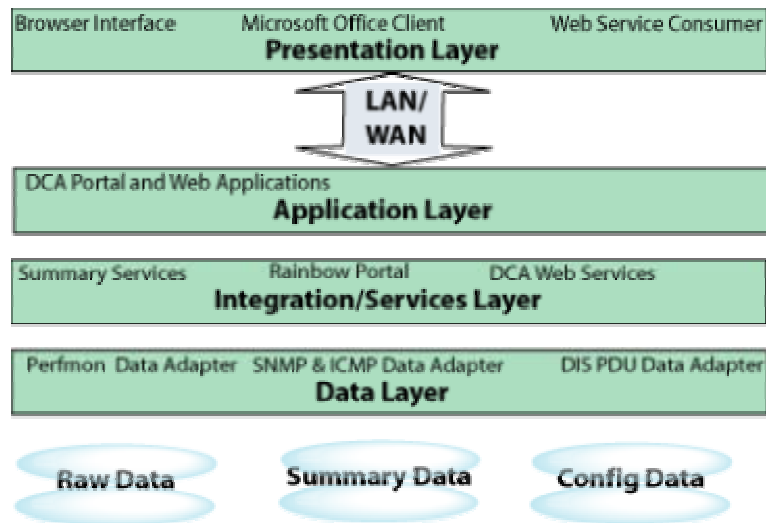


Figure 1: DCA Multi-layered service based architecture

The figure depicts four layers with services embedded in each layer. The figure captures the current state of the DCA Architecture. We have plans to expand each layer, for instance by adding additional Data Adapters to the Data Layer, but these expansions are not depicted.

The DCA has been designed to support distribution of key components. For instance, the Data Adapters can be and in fact were distributed to other machines to support the DCA effort on IL'03. The notion here is that it is easy to use different devices on the network to do data collection so that a) we can avoid overloading any given device and b) we can place bandwidth intensive Data Adapters on the same LAN (and even the same virtual LAN) as other important devices to minimize ill effects of highly distributed systems. Because the system is web-based, anyone anywhere on the network (connected via WAN or LAN and possessing the appropriate privileges) can log onto the DCA Portal (described below), run interactive queries against DCA databases and view, refresh, or publish analysis products.

The design has further been guided by the need for the Data Collection component to be real-time and for the analysis component to at least be near real-time. The Data Adapter framework incorporates real-time control and exception management. The multiple layers (e.g. separating the data layer from the integration/services layer) allow the analysis component to be able to yield limited results in real time or more complete results at near real time or after action.

Currently, a complete DCA system can be run on a single high-end laptop. As the load placed on the DCA system increases, either through the expansion of the collection requirements or through increased user traffic supported by the DCA Portal, additional servers are needed to distribute the load so that the Data Adaptors do not fall behind and the Portal remains responsive.

A guiding principle in the design of the DCA system is to separate functionality, implemented by code, from configuration. DCA collection parameters are typically specified in collection strategies, and system parameters are stored in name-value pairs, all of which is stored in a central database. The advantages of this approach are:

- Allows significant modification of DCA deployment without code changes.
- Non-developers can configure the DCA and the Data Adapters.
- Collection Strategies and system parameters can be configuration managed separately from the code base.

Currently, four XML collection strategy schemes have been develop, one for each Data Adapter and a fourth for the Device Discovery strategy which is used to specify address ranges to look for devices on the network. The hierarchical strategies specify which data to collect and at what sampling rate.

A web-based configuration module has been incorporated into the DCA Portal to allow DCA administrators to configure all aspects of a DCA deployment in a semi-automated fashion. Figure 2 illustrates the modules and summarizes the configuration process.

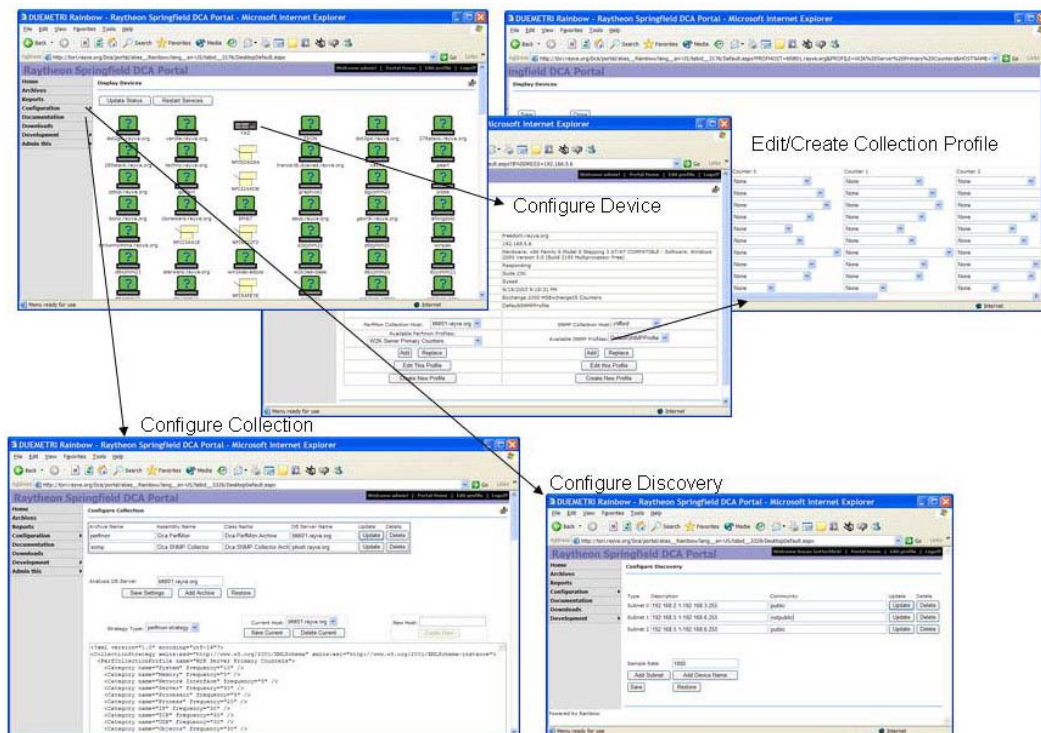


Figure 2: Automatic device discovery is used to semi-automated the configuration of collection strategies and other aspects of the DCA.

One feature of this architecture is its run-time extensibility. The Data Adaptors archive their collected data into a database which is typically on their local host. The DCA Portal can be configured to browse data from one or more hosts and one or more databases on each host, which is specifiable at run-time. An administrator merely needs to alter the

archives parameter through the DCA Portal configuration UI and reload relevant DCA Portal pages for the change to take effect. This approach allows for a completely distributable collection scheme which can be experimented with during deployment.

This flexibility, in part, is supported by our discriminate use of web services to support distribution across a WAN enterprise. There are many recent articles extolling the virtues of web service implementations, for instance see [1, 2]. Our DCA architecture utilizes web services both internal to itself and as an interface to other applications. We also utilize web query support, provided as part of recent Excel releases. This technology is similar to a web service only it supports a man-to-machine interface rather than a machine-to-machine interface.

The flexibility and extensibility of this architecture could easily be used to support similar applications such as Situational Awareness Displays or Collaborative Information Environments. It is our vision that ultimately such C2 applications will be developed on top of a multi-layered service oriented architecture and the DCA System will potentially be able to interoperate with these C2 applications on any layer.

### 3. DCA Portal

To facilitate easy navigation of DCA web-based capabilities, support publication of analysis products (reports, charts, etc.), and implement a user and role management system, our DCA system uses portal software as its baseline user interface. Currently, we have implemented the DCA Portal by extending the Rainbow Project's Portal, an open source initiative, to suit our requirements. The Rainbow Portal supports ASP.NET "best practices," is well supported and documented, and is extensible both at the installation script and the source code level. The Rainbow Portal incorporates many ASP.NET features that are needed now or in the future by the DCA System, including [3, 4]:

- Cross-browser support for Netscape and Internet Explorer
- Mobile device support for WAP/WML and Pocket Browser devices
- Clean code/html content separation using server controls
- Pages that are constructed from dynamically-loaded user controls
- Configurable output caching of portal page regions
- Multi-tier application architecture
- ADO.NET data access using SQL stored procedures
- Windows authentication - username/password in Active DS or NT SAM
- Forms authentication using a database for usernames/passwords. Encryption added for additional security so passwords are not stored in clear text
- Role-based security to control user access to portal content
- Support for existing modules for dynamic content
- Available in 14 (foreign) languages
- Allows content authoring to be safely delegated to **role-based** team members who need little or no knowledge of HTML

- Support of custom module creation for extensibility with 45 modules included in the standard release

Our customization of the Rainbow Portal occurs at both the installation script level and at the custom module level. By modifying the installation script, we automatically load our DCA specific content into the portal. By adding custom components we extend the portal to include DCA specific capabilities such as interactive database querying and analysis product publishing.

Note that the built in support for WAP/WML enabled devices provides a key benefit to the military user. Our use of Microsoft Excel as our central analysis tool further supports this. A “disadvantaged” user, who is using a “disadvantaged” device such as a PDA running Windows CE and Office CE, or using a “disadvantaged” network connection such as a surface-to-air link, can download an Excel file containing the raw data of an analysis product and render the report locally.

The DCA Portal supports user-initiated web-based customization of each of the tabbed panels, where customization access is restricted via the definition of four user roles:

- DCA Developer (Level 1 User) – The overall look and feel of the DCA Portal for a given deployment will be controlled by a group of individuals, for convenience referred to as DCA Developers. Essentially, this group is chartered with the task of installing and maintaining the DCA System according to a given distribution scheme and deployment strategy. The portal customization is viewed as part of the installation process.
- DCA Administrator (Level 2 User) – Various aspects of the DCA Portal, including some of the content customization, will be performed by a trusted group of individuals referred to as DCA Administrators. This group is chartered with the task of administering the DCA System and will perform activities such as: user role management, database administration and backup, and strategy file definition and maintenance. This group implicitly includes all DCA Developers.
- DCA Analyst (Level 3 User) – This is a group of individuals who has the necessary training and skills to develop charts, reports, and other DCA Analysis products. This group will be chartered with the population of the Charting tabbed panel and Reporting tabbed panel that the.
- DCA User (Level 4 User) – This group of individuals consists of the end users of the DCA and will have little, if any, portal customization privileges.

Based on the deployment strategy, subgroups of these four groups may be established. For instance, J5 staff may only be able access J5 content and J4 staff may only be able to access J4 content. Thus, a J5 DCA Analyst subgroup and J5 DCA User subgroup may be established. An image of the DCA Portal is shown in Figure 3. Several custom modules have been built to support DCA Configuration, interactive query, report (analysis product) publishing and scheduling, and live report updates.

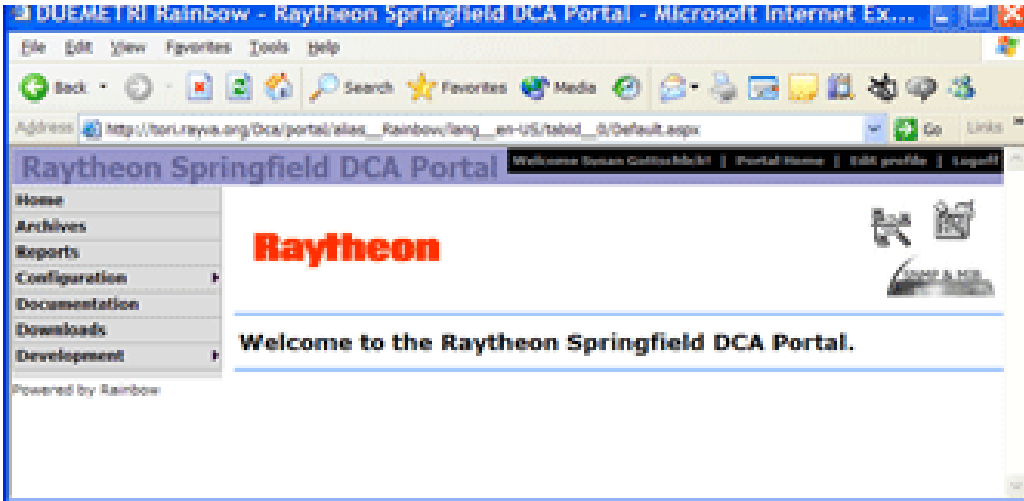


Figure 3: The DCA Portal uses basic portal tab infrastructure to support easy navigation through the web site.

Figure 4 illustrates the Archives module which allows the user to develop and run interactive web queries against DCA Collection Databases.

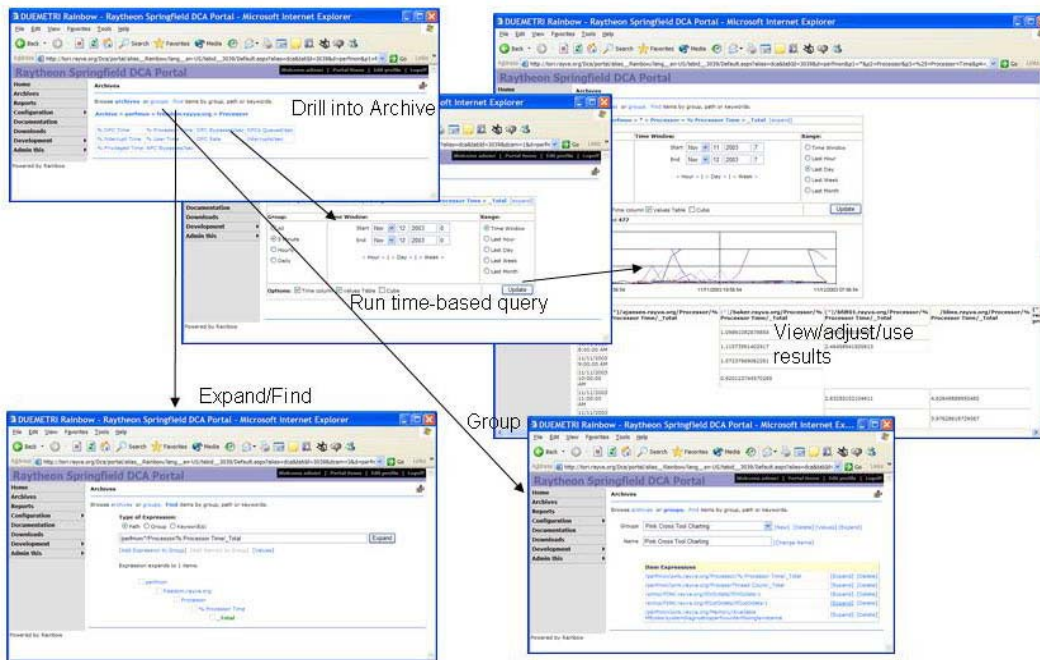


Figure 4: The Archives module lets a DCA Analyst drill into the available archives to develop queries, allowing the analyst to use Find/Expand for archive search and query editing and allowing the analyst to create groups of queries. Query results are presented in user-selectable chart, tabular, and/or cube form.

Reports (analysis products) in our DCA system are Excel workbooks with embedded web queries. DCA Analysts use Excel's Web Query interface, which in turn walks the analyst



through the Archives module, to extract and embed raw data into an Excel report. They then use Excel calculations and macros to process the raw data, create meaningful titles and labels, and ultimately create a visual product such as a strip chart, pie chart, pivot table, histogram, or data table. It is this visual product that is rendered to the web browser when the report is “viewed”.

Figure 5 illustrates the web module used to publish, modify, and view reports. A DCA User can schedule a report to be regenerated (web queries updated) for a given time period. A scheduler service automatically generates the updated reports and publishes them via the report history browser so any user can walk through and view reports that have been generated. Alternatively, the user can view “live charts” – i.e. reports that are refreshed and redisplayed frequently or with the click of a button.

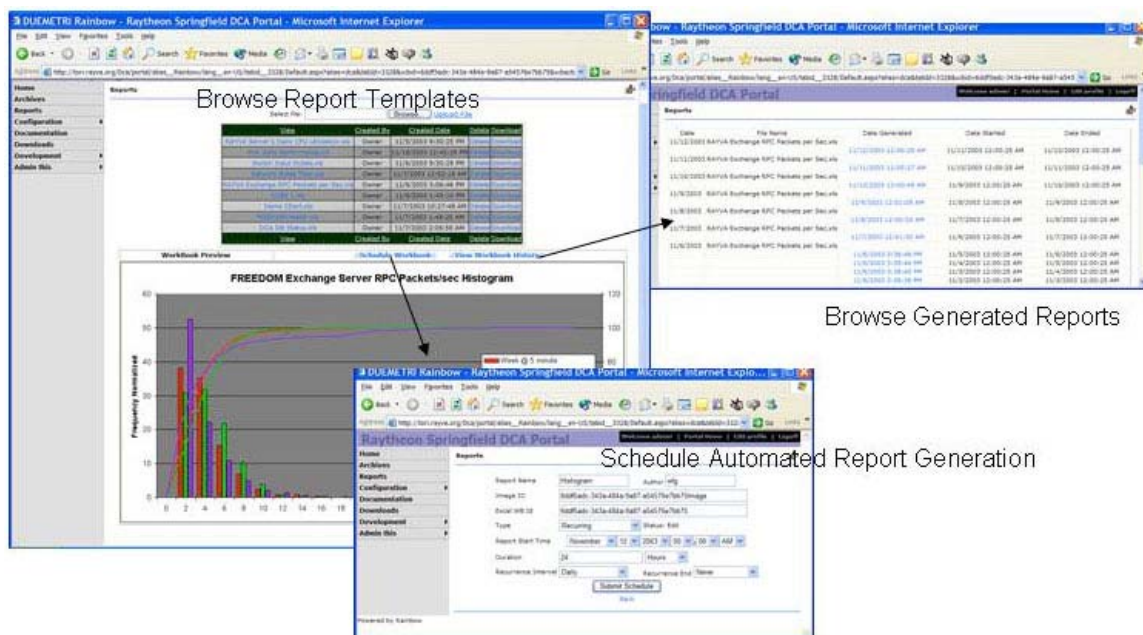


Figure 5: The Reports module lets a DCA Analyst publish a report or lets a DCA User download a report. A user may also schedule a report which will generate a new version of the chart for a specified time period (different from the original report). Once generated, the new reports will show up in the history browser. Alternatively, the user can view a live chart that gets updated frequently or with the click of the Update button.

We have found that given Excel’s powerful charting and data manipulation capabilities and its extensive usability features that DCA Analysts, with very little training, can produce very complicated products quickly.

#### 4. Summary and Future Work

We have presented a DCA system which is based on a multi-layered, service oriented architecture and is designed to be flexible and extensible. Design decisions have been motivated by the requirements levied by complicated distributed collaborative military

workflow processes. Heavy use of interoperable standards, COTS products and open systems software have yielded an extremely powerful system requiring a minimal development effort.

Future work can be divided into three main categories: addition of new Data Adapters, extending analysis capabilities, and miscellaneous improvements.

The design and implementation of additional Data Adapters will be driven by target applications. For instance, Data Adapters for US/MTF message traffic, HLA message traffic, and related message traffic, is anticipated.

Currently our analysis products are based on Excel workbooks. We would like to develop alternative interfaces that allow users to create and view analysis products based on other client applications as well.

Finally, we would like to continue making usability improvements, provide better “mobile” support for disadvantaged users, and develop a rich set of sample analysis products that can be quickly and easily modified to support the needs of new DCA deployments.

## **5. References**

- [1] Sutor, Bob (2003), “Plumbing Web Connections,” *Harvard Business Review*, 81(9), 18-19.
- [2] Special Issue on Web Services Computing, *IEEE Computer Magazine*, October 2003.
- [3] Vertigo Software Inc. (2002), “IBuySpy Portal: Design and Implementation,” IBuySpy White Papers, Microsoft Corp., January 2002, 1-19.
- [4] Dumetri Rainbow, <http://www.rainbowportal.net/>