Homeland Emergency Response Exchange (HERE)

(Previously known as: Heartland Emergency Response Exchange)

Lessons Learned

Final Version

January, 2008

WINDSOR
SOLUTIONS, INC.
Environmental + Health
Information Systems

THIS PAGE INTENTIONALLY LEFT BLANK

Contents

THIS PAGE INTENTIONALLY LEFT BLANK

Introduction

Purpose

The Homeland Emergency Response Exchange (HERE) project was performed in 2006 through 2007, and resulted in the HERE Network (www.HERENetwork.org). This Network allows emergency personnel to gain access to pertinent and valuable data from environmental agencies.

With the initial development and rollout of the Network complete, this document examines some of the key issues that emerged during the project and the lessons gained though their assessment and resolution.

This document is intended to provide useful insights for future HERE implementers, as well as other Exchange Network flow developers as they consider their architectural options for future, similarly innovative data exchanges.

Lessons Learned

Reuse of Exchange Network Products

Situation

One of the critical success factors of the Exchange Network (EN) is for reuse of products versus recreation or replication. This is important not just for efficiency, but to avoid contradictory and confusing alternatives. The EN has employed a variety of mechanisms to support this principle (e.g., the Core Reference Model and Shared Schema Components - SSCs), but the HERE project employed a novel variation of such reuse.

Emergency responders have a need for a variety of environmental data, but they rarely need all of the detailed data that the environmental agency manages. Existing EN flows support the sharing of those same (but more expansive) datasets. One example of this is for details of the Public Water System (PWS); environmental agency needs call for a large variety of data whereas emergency responders only value a portion of that data.

There are a couple of ways in which reuse could be applied to this type of situation:

1. Create new custom XML schema for emergency response needs, and reuse the common shared components for those data elements that are of mutual need, while excluding those components that are of no interest to emergency responders.
2. Reuse an entire schema and assume that the exchange partner will disregard the data that they are unconcerned with.

For HERE, there were four sets of priority data: Facility Identification, Livestock, PWS, and Chemical Inventory. For each of these there are existing EN flows already in place. The latter approach was chosen for the following reasons:

- Some of these existing flows had already been implemented by the States, and so reusing those flows would mean a reduced cost of implementation and ongoing maintenance. The PWS exchange is an extreme example of this. Some states have implemented the SDWIS data flow (using FedRep to create those data files). By adopting the SDWIS Inventory XML Schema for HERE, very little work was required to exchange this data for HERE.

- For flows that do not already exist at an agency; by developing HERE flows that share data using the existing schema, an agency will be far along the road should they wish to implement the full version of that flow at a later time. For example, a couple of states implemented the CAFO XML Schema to provide Livestock data. Should another Partner now wish for the full CAFO exchange to be implemented at those States, they will have a much reduced implementation effort.

Conclusion

With the project now completed, these decisions can be assessed retrospectively. The advantages identified have remained valid, and no unforeseen issues occurred during implementation. This approach is recommended for other EN projects that intend to exchange *large subsets* of data that are already contained within a preexisting exchange.

High Volume, High Frequency, Multi-Recipient Data Exchange

Situation

Most existing Network Exchanges have required data to be exchanged with one recipient Partner. Even in these instances, the production and processing of large XML files has had some challenges. When the target audience is potentially thousands of data recipients, and requires the exchange of large volumes of data, then these challenges are far greater.

The HERE Network was developed to support a variety of emergency responders, and although the breadth of its adoption is still unclear, this may be a very large audience. Furthermore, the volume of data exchanged is large (including full sets of data from four traditional flows). The data needs to be kept as up to date as possible, and so the frequency of exchange is also high.

This issue was recognized early in the project and a variety of approaches considered. One extreme solution would be to create large full-replace files for each flow every night and thus always provide the latest data. However, that would then put a massive bandwidth and data processing strain on the receiving Client, which is typically running on a commonplace personal computer.

Therefore, small, incremental transmission of changes to the data were necessary to ensure that the Clients could be synchronized without their PCs being strained. This though, would require that the Nodes produce incremental files for each Client, and if the number of Clients grew, could also cause a scaling issue for the Nodes.

The compromise approach chosen was to produce daily incremental files for each dataset, and then allow those pre-prepared files to be downloaded to each Client. So only four small files need to be generated each night by a Node, and each Client performs a simple download of all such incremental files since the Client's last update, and processes this set of files. This provides the least processing demands on both Node and Client.

Conclusion

The expansion of the HERE Client has yet to occur (having just been released), so the scalability has yet to be proven or challenged. However, during testing, the incremental approach was proven to be highly efficient for both Node and Client. This approach does add a level of complexity to data preparation and collection, and required the provisioning Nodes to support a style of operation that is probably unique to this flow. Development time definitely increased, but the anticipated issues that would have occurred using more traditional Node exchange should be averted.

Network Stability and Resilience

Situation

The Exchange Network is providing a medium for partners to gain access to data that has historically been hard or impossible to attain. As this continues, some Partners become reliant on the data that they now have immediate access to. This then demands far greater assurance that the data will be available at all times. One solution that has been attempted in other EN projects has been to employ a form of data source redundancy; such as a central server to replicate the data using the EN as the conduit.

However, emergency response presents further challenges to these solutions. First, data access needs become critical during emergency events, and there can be no delay in obtaining the necessary data during those events. Second, the Internet may be inaccessible due to the location of responders in the field or damage to the network.

These needs prompted the HERE team to employ a new approach to support data exchange. The data needs to be accessible on a laptop for remote access, and also needs to be cached locally so that no network stability or performance issues limit its usefulness. To achieve this, a local data store was required, and a data collection and aggregation process was needed that could process the data on a personal computer with limited power.

Conclusion

Data caching has the disadvantage that the data can become stale; but the data that HERE exchanges has a fairly low volatility, with few changes occurring on a daily basis. This low volatility allowed a local caching and incremental update approach to be used.

The benefits of data caching were immediately obvious. There were occasions during the development of the exchange where the HERE Client was used to demonstrate the types of data that could be accessed using the tool. In those situations it was occasionally the case that there was no Internet access available—in some cases due to the location, but in others due to Internet access failure. The HERE Client operated readily even without direct access to the Network Nodes that supplied the data. Furthermore, application performance was predictable as a personal computer does not suffer the same risk of concurrent resource demands as does a multi-user server.

Data Sensitivity and Authorization

Situation

Some of the data that is made available through the HERE exchange is considered highly sensitive as it provides information that could be used with malicious intent. The Exchange Network provides a reasonably secure mode of data transportation but it uses newer technologies and presents some mystique to those not intimate with the Network.

More traditional approaches to securing data access have typically been employed by the stewards of the HERE-sensitive data, such as disallowing the data to be disseminated outside of the agency, or only providing data on CD via registered mail. This has resulted in the data that is infrequently updated or inaccessible to those who are valid consumers as well as those who are not.

To address these needs in the HERE Exchange, a few prerequisites for the exchange were identified:

1. Only authorized users can access the HERE data, and they must have agreed to the conditions of use as specified in the Trading Partner Agreement.

2. Some datasets are much more sensitive than others (at one extreme, some datasets are available on public web sites already). So authorization must allow for selecting dataset-level granularity. For example, one user may have access to the drinking water data and the livestock data, but another may only have access to the livestock data.
3. Users will need to attain data from many Nodes (a goal of this exchange is for emergency responders to have access to data that crosses jurisdictional boundaries, a reflection that disasters often do not respect such boundaries). Administration of such authorization should therefore be centrally managed to reduce the user's burden of remembering multiple sets of credentials.

NAAS is the clear choice for such central authorization. However, EPA's NAAS management utility is limited in its ability to administer flow level user authentication. This can be administered using the NAAS web methods; but the NAAS Administration user interface does not support this granularity. Furthermore, some legacy Nodes do not have the ability to validate authorization from NAAS to this level—in fact, many Nodes allow access to all their data services to any user that has authority to access their Node.

Conclusion

The policy issues took time to overcome. A fair amount of education and reassurance was required so that the stewards of sensitive data accepted the exchange of their data. Ultimately, the resolution of the other states and the intent of this collaborative initiative provided a good incentive to share.

Technically, the exchange demanded the deployment of Node enhancements to support the requisite flow level authorization via NAAS. A basic authorization request process was also put in place (via email), but with the expectation that if the number of HERE users grew significantly, it would justify a more elegant central administration alternative. Such a central server would coordinate the multi-node authorization for new users, i.e.:

1. Register new user's requests for access from each HERE Node and for each selected HERE dataset.
2. Allow HERE Administrators to create new NAAS user accounts.
3. Request authorization from each Node Administrator for access to the datasets requested by the new user.
4. Notify the user as each of these authorizations are approved.

Central HERE Administration

Situation

As mentioned above, a central HERE Administration Server would provide a more robust solution for the oversight of the HERE Network, should it continue to grow as is predicted. The server would provide a more elegant solution for managing the user authorization process, but it might provide some additional capabilities that will ease the ongoing maintenance of the exchange:

1. Manage the expansion of the HERE Network:
    a. Provide a central location to inform of upcoming changes to the Network.
    b. As new Nodes and/or flows are added, notify users so that they can request additional data access if they so wish.
2. Support the HERE Client application:

a. Provide a location where the Client can subscribe to software updates.

b. Provide support information such as troubleshooting advice and a user forum.

3. Support the administration of common HERE data and documentation:

a. The HERE Client relies of a variety of contextual information to help users interpret the data presented. This contextual information may need to change over time. For example, to support additional or updated chemical reference information.

b. Provide an interface to allow easy additions to the contextual information. This updates would then be distributed automatically to the Clients.

Conclusion

With the current size of the Network limited to early adopters, these administration tasks are not overly-burdensome. The project team has already developed some plans to support these additional capabilities and intends to incorporate them if HERE expands as is anticipated and funding can be appropriated.

Deployment issues

Situation

A PC-based HERE Client was chosen as the user interface over a web-based Client for a variety of key reasons, including:

- A desktop or laptop PC's ability to retain a local cache of data and thus be network independent
- The distribution of data processing demands, which can be considerable during an emergency
- The variety of powerful third-party mapping capabilities that can be installed on a PC
- The fact that users may be using HERE only as a way to load this data into their own backend databases, and desktop software has much greater ability to support such functions.

However, deployment of PC-based software has its challenges. The HERE Client was developed to work on the majority of today's computers, without unrealistic requirements. It includes the ability to monitor for software updates and apply those changes in place. It relies on SQL Server Express for its data store which is a free RDBMS.

During development and testing, deployment issues were at their extreme, with regular re-releases of the software and variations in environment being exposed. By using a fairly diverse set of Beta testers, many installation issues were identified and resolved for future installations.

The greatest issue experienced was the variations in which the agencies access external sources. Depending on the specific Internet connectivity required for that agency, the client had to be flexible enough to sense the user's network and apply the necessary network/proxy configuration along with the authentication scheme appropriate for that specific environment.

Considering that individual clients are often deployed on mobile tablets and/or laptops the HERE client had to be also "smart" enough to sense a new environment and adapt itself either automatically or, when not possible, ask the user for details to aid that configuration process.

Conclusion

Significant issues were encountered during the testing phases of the project and accordingly resolved. Since release to production no new installation issues have been encountered, but it is too early to be sure that such issues will be as nominal as anticipated.

Facility/site reconciliation

Situation

One of the assumptions made early on about the data exchange was that users should not have to suffer interpreting replicate data coming from the multiple datasets that HERE supports. Environmental agencies tend to collect data because of various regulatory reporting requirements and within various data management systems. This can result in duplicate information about the sites or facilities that are of interest to emergency responders.

The Facility Identification flow provides a mechanism to transmit reconciled sets of site data to the HERE Client; and so it was decided that this was a necessary medium. However, during flow implementation, it was found that some datasets were not yet integrated and reconciled within the environmental agency, and it was unlikely that they could be reconciled in time.

Conclusion

Emergency responders will find the HERE data easier to interpret if that data is reconciled. However, the duplication of some sites does not make the information unusable, just more cumbersome. Therefore, for the Chemical Inventory dataset, the Client aggregates facility data from this flow along with the Facility Identification dataset and presents the combined data to the user seamlessly. This capability may be extended to the other datasets should the need arise.

For the PWS data (using the SDWIS XML Schema), the ‘site’ data has far less overlap with the other environmental data. It is relatively rare that a drinking water well, intake, or treatment plant is also associated with other environmental activities such as storing hazardous chemicals onsite. Furthermore, given that many agencies use the pre-packaged SDWIS/FEDREP XML files, this is a good candidate for the same approach.