Avian Flu Case Study with nSpace and GeoTime

Pascale Proulx, Sumeet Tandon, Adam Bodnar, David Schroh, Robert Harper, and William Wright*

Oculus Info Inc.

ABSTRACT

GeoTime and nSpace are new analysis tools that provide innovative visual analytic capabilities. This paper uses an epidemiology analysis scenario to illustrate and discuss these new investigative methods and techniques. In addition, this case study is an exploration and demonstration of the analytical synergy achieved by combining GeoTime's geo-temporal analysis capabilities, with the rapid information triage, scanning and sense-making provided by nSpace.

A fictional analyst works through the scenario from the initial brainstorming through to a final collaboration and report. With the efficient knowledge acquisition and insights into large amounts of documents, there is more time for the analyst to reason about the problem and imagine ways to mitigate threats. The use of both nSpace and GeoTime initiated a synergistic exchange of ideas, where hypotheses generated in either software tool could be cross-referenced, refuted, and supported by the other tool.

CR Categories: H.5.2 [Information Interfaces & Presentations]: User Interfaces – Graphical User Interfaces (GUI); I.3.6 [Methodology and Techniques]: Interaction Techniques.

Keywords: visual analytics, information visualization, human information interaction, sense making, geo-spatial information systems, temporal analysis, user centered design

1 INTRODUCTION

GeoTime and nSpace are two novel visual analytic applications that have been developed in collaboration with analysts to support the investigation of large and complex datasets. nSpace is used for triaging massive data and for analytical sense-making [12], [22], and is currently undergoing experimental evaluation and pilot deployment. GeoTime supports the visualization and analysis of entities and events over time and geography [13] and is currently in transition to deployment for analysts to use on a day-to-day basis.

The epidemiology dataset was created by the authors' for the purposes of investigating the analytic process and evaluating nSpace and GeoTime using an ecologically valid scenario. This dataset is based on real news reports detailing the current outbreak of avian flu. The analyst used for the evaluation was experienced with both nSpace and GeoTime, but had no formal training in the art of intelligence analysis.

* email: {pascale.proulx, sumeet.tandon, adam.bodnar, david.schroh, rob.harper, bill.wright}@oculusinfo.com

1.1 nSpace - A Unified Analytical Workspace

nSpace combines several interactive visualization techniques to create a unified workspace that supports the analytic process. One technique, called TRIST (The Rapid Information Scanning Tool), uses multiple linked views to support rapid and efficient scanning and triaging of thousands of search results in one display. The other technique, called the Sandbox, supports both ad-hoc and formal sense making within a flexible free thinking environment. Additionally, nSpace makes use of multiple advanced computational linguistic functions using a web services interface and protocol [6, 3].

1.1.1 TRIST

TRIST, as seen in Figure 1, uses advanced information retrieval methods to support rapid scanning and exploration of large datasets. The design of TRIST was informed by recent research into information retrieval [4,7,17], the development of the cognitive dimensions framework [5], and the idea of an information-seeking environment [8].

TRIST supports multiple analytical processes, including query planning, rapid scanning of large corpus of documents, and result characterization across multiple dimensions [12]. Analysts work with TRIST to triage their massive data and to extract information into the Sandbox evidence marshalling environment.

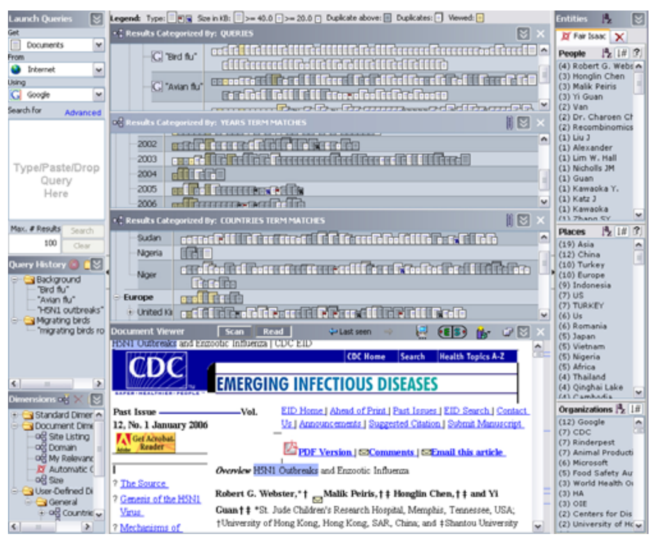


Figure 1: TRIST interface. Left column: Launch Queries, Query History, and Dimensions panes. Middle column: Displayed dimensions with categorized results and Document Viewer. Right column: Entities pane.

An evaluation of TRIST conducted over a two day period by The National Institute of Standards and Technology (NIST), found that analyst productivity and quality of analysis increased when using the tool [18].

1.1.2 Sandbox

The Sandbox is a flexible and expressive environment for working with evidence and generating hypotheses. While concept maps [11,15] and link analysis tools [10,19] aim to support similar activates, their node-link representations can be limiting, and readability and usability decrease as the volume of data increases.

The Sandbox supports both ad-hoc and more formal analytical sense making though the use of fluid gestures for interaction. Additionally, the Sandbox supports automated process model templates, layers for managing multiple dimensions of data, assertions with evidence templates, and scalability mechanisms to support larger analytic tasks [22].

An evaluation of the Sandbox was conducted by the NIST Visualization and Usability Group in 2005 [16]. Four analysts participated in the experiment consisting of three hours of training and eight hours of work on an intelligence analysis task. Results indicated that the Sandbox enabled the analysts to work quickly and efficiently, and perform a higher level of analysis when compared to other tools.

1.2 GeoTime - Concurrent Temporal and Spatial Analysis

GeoTime improves perception and understanding of entity movements, events, relationships, and interactions over time within a geospatial context. As seen in Figure 2, events are represented within an X,Y,T coordinate space, in which the X,Y plane represents geographic space, and the T axis represents time. In addition to providing spatial context, the ground plane marks the instant of focus between past and future, meaning that events along the timeline 'occur' when they meet the surface.

Several other visualization techniques for analyzing complex event interactions only display information along a single dimension, typically one of time, geography or network connectivity.

An evaluation of GeoTime found that its unified geo-temporal representation increased analysts' understanding of entity relationships and behaviors [13].

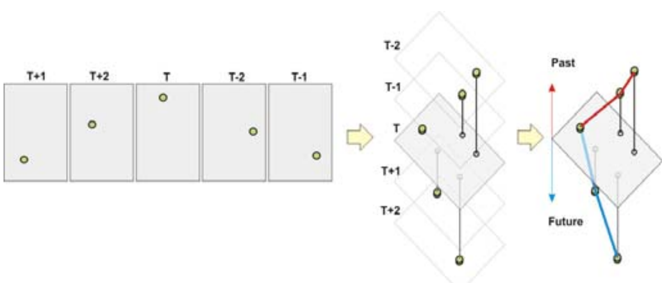


Figure 2: Individual views of movement in time are translated into a continuous spatiotemporal representation.

1.3 Epidemiology Analysis

Epidemiology is the study of the distribution and determinants of health and diseases, morbidity, injuries, disability and mortality in populations [14]. In this context, an epidemiologist can be thought of as a disease detective. Common tasks involved in epidemiology include surveillance, outbreak investigation, control measures, risk assessment, event characterization, and long term impact research and the evaluation of interventions. Trend analysis and change detection are also important in epidemiological analysis. Changes can occur in disease occurrence and distribution, in agent or host factors, or even in health care practices. Typical analysis tools used in epidemiology include statistical measures, spatial analysis, geographic information systems, and time series analysis tools [1].

2 THE AVIAN FLU CASE STUDY

The avian flu case study presents several analytical challenges. The dynamic nature of the data requires tools which support temporal analysis, and enable the understanding of complex geo-spatial events. The objectives for the case study are listed in Table 1. The role of the analyst will be played by *Lynn*, a fictional analyst who is familiar with both nSpace and GeoTime, but has no formal training in analysis, and can thus be considered a novice user. The case study will provide a detailed examination of the analytic process as Lynn uses nSpace and GeoTime to fulfill the objectives of the exercise. The data used in this case study is publicly available and consists entirely of internet sources obtained through TRIST via the Google search engine. Key references include the World Health Organization's avian flu disease outbreak news [21] and events timeline [20].

Table 1. Analysis Objectives

- Investigate and characterize the avian flu outbreaks
- Identify and analyze trends
- Identify key issues and threats
- Develop, support or refute, and refine hypotheses about cause and effects
- Propose options for mitigating future risk

2.1 Thinking without Borders

Lynn has just received the assignment to analyze the distribution, determinants, and potential impacts of the recent avian flu outbreak. She begins by brainstorming in the Sandbox using her prior and tacit knowledge to generate hypotheses about the possible causes and effects of the recent outbreaks. She notes key questions and outlines her analytical strategy for further investigation. The fluid gestural interaction within the Sandbox supports the early stages of analysis, which are often unconstrained to enable the generation of new ideas and concepts. Lynn simply points and clicks to create annotations. She draws circles to create groups and draw links to connect related ideas. Using these techniques, Lynn's thoughts flow freely, and are quickly organized with drag and drop and semi-automatic layout. Emphasis can also be added to create visual structure. Alternative scenarios are sketched out and the resulting empty assertions become explicit reminders to find supporting and refuting evidence, as shown in Figure 3.

Figure 3: Brainstorming in the Sandbox.

2.2 Acquiring Analytical Awareness

With several avenues to investigate, Lynn turns to TRIST where she can follow a deductive process and find answers to questions, as well as and evidence to evaluate each of her assertions and hypotheses. As nSpace does not restrict the analyst to a specific workflow, Lynn is able to adjust her analytic process based on her current task. Rather that continue her analysis with pointed queries, Lynn decides to use TRIST to immerse herself in the data and favor a more inductive type of reasoning. Lynn's goal is to avoid or delay time investment in a particular scenario, which has the potential of introducing analytic bias.

At this point, Lynn could point TRIST exclusively to the Health Department internal database, but believes that could also introduce some bias due to its narrow focus. At this early stage, Lynn might miss some signs of new outbreaks or key contextual cues available on the web. The web however poses a significant challenge to the epidemiologist attempting to perform public health surveillance. The signal to noise ratio is very high in this particular case. The potential gravity of the avian flu threat and its global reach generates a massive amount of information, in constant flux, from sources of variable reliability. TRIST allows her to evaluate the relevance and source credibility of thousands of search results at a time, by revealing various attributes and content without having to open and read results.

Lynn opts for this iterative analytical technique, rather than relying only on reports sent to the Health Department. Starting with general queries that she refines iteratively, Lynn uses the TRIST launch queries pane, to examine documents from the web using Google with the query "Bird flu". Rather than overwhelm Lynn with the several million query results, TRIST presents the first 100 documents. Instead of starting with this result set, Lynn decides to launch a few variations such as "Avian flu", and "H5N1 outbreaks". Using TRIST's difference visualization capabilities, she can quickly compare and determine what is unique or common within each result set without having to open a single document. An example is shown in Figure 4.

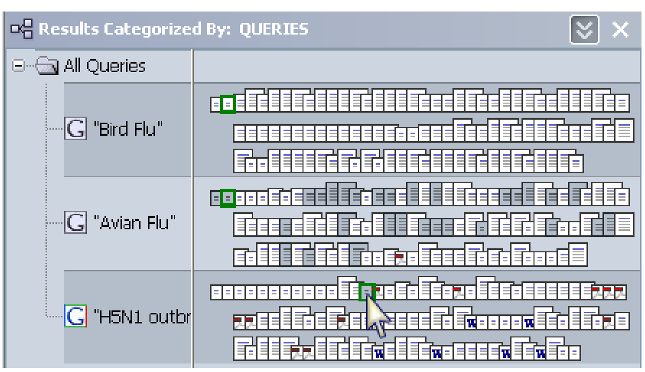


Figure 4: TRIST difference visualization. Grey icons indicate previously retrieved results, while white ones are new results.

Using this process, Lynn can close the feedback loop on the query refinement process while increasing the precision and recall of her information retrieval process.

Looking in the entities pane, as seen in Figure 5, Lynn discovers that there seems to be a significant number of documents on particular countries that are affected by avian flu. She drags these entities into the sandbox and unrolls them into a group so she can add information on the current situation in each entity group later on. As she clicks on the entities in the sandbox, all the documents that contain each entity become visible, and the entity terms are highlighted in green.

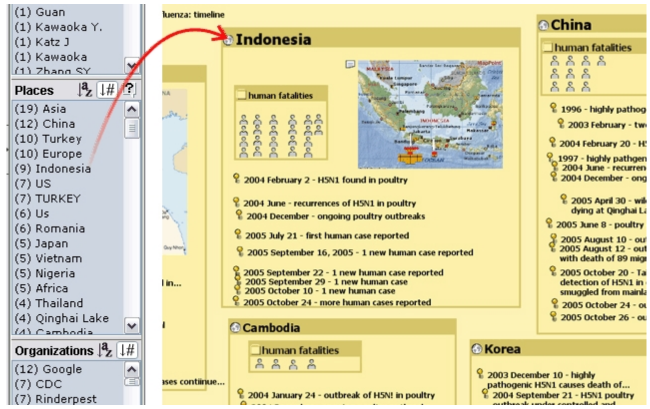


Figure 5: Extracted entities can be used to browse results or help determine relevance. Entities of interest are dragged to the Sandbox where they stay linked to results containing that entity.

2.3 Discovering Trends and Themes

To establish the existence of an outbreak, categorized as a sudden increase in the incidence of disease, Lynn performs a rough trend analysis using the Country and Time dimensions in TRIST, which automatically sort all query results into each dimension. Without opening a search result, the categorization hints at various observations that can be investigated further at a later time. For example, Lynn notices that the disease was first identified in Hong Kong in 1997 and that most human fatalities seem to have occurred in Asia and in the Middle East. Considering a query result as a unit of incidence, the length of each row in each dimension bin reveals some trends. More observations are available by selecting rows in one dimension and analyzing the highlighted results in a different dimension, as shown in Figure 6. For example, Lynn can focus on the evolution of the disease per country or get a sense of the propagation of the disease over time.

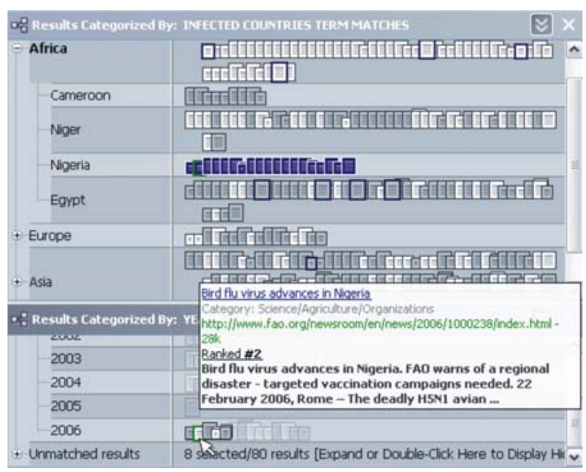


Figure 6: Results from multiples queries are automatically organized in any dimension of interest created by the analyst. Selection-highlighting across dimensions reveals trends without requiring the analyst to open a single document. In the Country dimension, Nigeria results are selected in blue and those same results are highlighted in the '2006' bin of the Time dimension.

Lynn captures these quick observations, either as snippets, whole documents, or sets of documents, in the Sandbox. She notices that the World Heath Organization's website was recurring across TRIST's country and time dimensions. This indicates the availability of comprehensive data within a narrow source, and on closer inspection, Lynn sees that it contains spatiotemporal data of interest.

Before moving on to examine that spatiotemporal data, Lynn opens the Automatic Categories dimension to see if any novel or unexpected themes are detected in the data. The default generation of three themes doesn't show her anything new, so she asks for four more categories to be generated. One of those, 'animals' seems potentially interesting, so she asks that it be further broken down into four categories. The results come back as 'chickens', 'turkeys', 'cats' and 'swans' (Figure 7). Lynn is reminded that while she tends to associate avian flu primarily with domesticated birds (probably because of a large number of recent news stories about that issue), it also affects wild birds, which could certainly affect the chances of it being contained. She was also interested to see cats detected as a theme in quite a few of the results, since she hasn't heard much about outbreaks in cats. She wonders how much of an affect the disease is having on this non-avian, mammalian population. She'll keep these issues in mind as she continues her analysis.

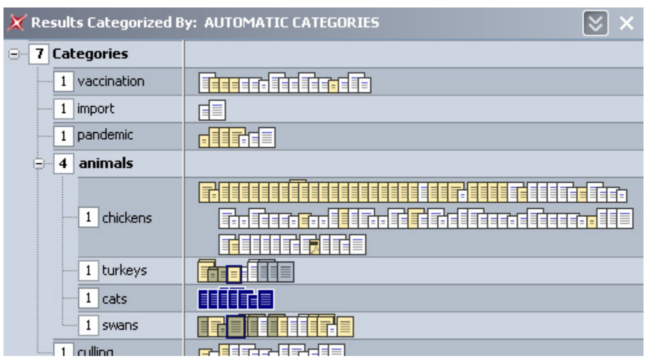


Figure 7: Themes contained within the results, located by and displayed in the Automatic Categories dimension.

Having gained a better overview of the situation, as well as some hypotheses regarding trends and issues, Lynn turns to GeoTime to gain a different perspective on the scenario.

2.4 Disambiguating Spatial and Temporal Patterns

At the core of GeoTime is the concept of an event, which is loosely based on Davidsonian semantics [2]. An event binds an action or occurrence to a location, a time, and the entity or target involved. It links the what, the where, the when, and the who to a single entity. Each event can have a description or type, a location, and a temporal context.

Lynn structures the avian flu data within GeoTime as follows: (1) each event represents one or more human cases of disease within a local area, labeled by age/sex of the victim and, in the case of a multi-victim outbreak, the number of people infected, (2) each event is associated with the strain of avian flu involved and those strains are the 'targets' in this scenario, and (3) other potentially relevant information, such as method of exposure, is saved with each event. To import the data into GeoTime, an Excel plug-in is used which translates tabular data into places, events, and targets in GeoTime.

When working with data sets in which the targets represent actual moving entities, such as vehicles, shipments, or in this case, a strain of flu, the target trail represents the targets movement, and depending on the accuracy and granularity of the data, can be fairly accurate and used for analysis.

In this case, each target represents a disease strain, which we know can exist in multiple discrete populations at the same time and move in more than one direction at once. Rapid movements of this target are therefore indicative of simultaneous and widespread occurrences in an area encompassed by the visible target trails. These trails thus provide an idea of the magnitude of global disease activity for a particular strain. Removing the trails from view gives the analyst a view only of individual disease events.

With the dataset imported into GeoTime, geo-spatial and temporal patterns are immediately evident, as shown in Figure 8.

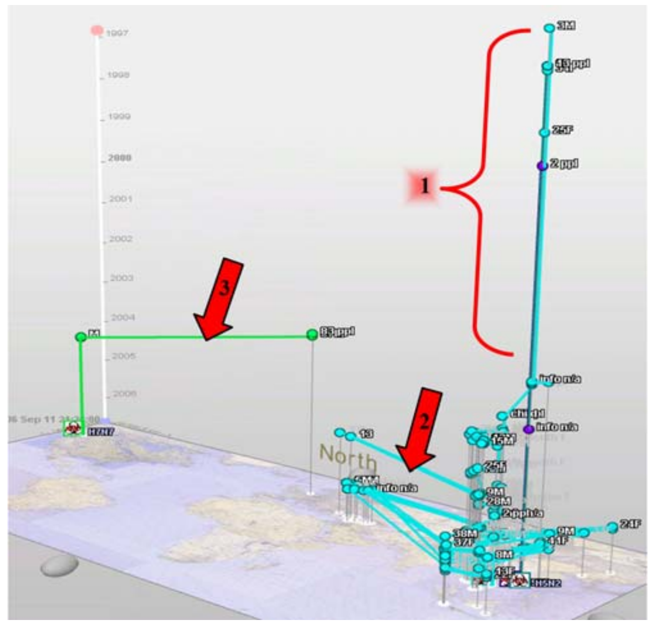


Figure 8: Initial view of the data, showing worldwide human cases of avian flu. Sporadic localized activity seen in Asia until 2003 (1). H5N1 cases increased in frequency in 2004 (2) with the bulk of disease activity in Asia. Isolated incidents of strain H7N7 found in Canada and Europe (3).

Geographically, the bulk of disease activity is clustered in Asia, with some occurrences in the Middle East and a few in Northern Europe and Western Canada. Temporally, disease activity appears sporadic between 1997 and 2003. Target trails, one per strain, indicate substantial increases in disease activity in 2004, continuing through the end of the dataset in early 2006. A particular strain, H5N1, is clearly predominant, and is present in Asia and the Middle East. Africa has no reported cases of avian flu, while Europe and the Western Hemisphere appear to have each had a brief and isolated incident with a strain not seen elsewhere.

Using GeoTime, Lynn develops hypotheses about why disease events and patterns occur. These hypotheses require further investigation to assert or refute their validity, but the generation of these hypotheses is necessary in understanding the spread of disease on a global scale.

Since Asia is a hotspot for disease activity, Lynn decides to investigate this region more closely, as seen in Figure 9. Lynn notices that the concentration of three different strains, shown as three distinct colors, in this geographic area is apparent, which may imply that mutations have taken place within this region. Figure 9 also reveals that disease events in Asia seem to occur in clusters by location, with time gaps in between outbreaks. The absence of the disease from nearby countries like Laos and Korea are also interesting observations. Indonesia is not connected to mainland Asia, yet somehow disease events later occur there, suggesting avian flu was somehow brought over from the mainland, perhaps due to bird migration or poultry exports. Lynn quickly creates new assertions in the Sandbox to capture these thoughts and drags a thumbnail of this GeoTime view to place next to them so that she can quickly jump back to it from the Sandbox in the future, as seen in Figure 10. Clicking on a GeoTime thumbnail in the Sandbox invokes that context in GeoTime.

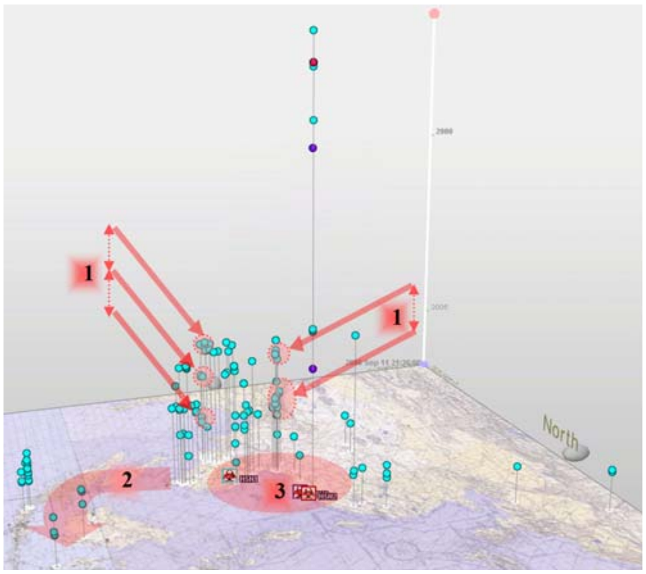


Figure 9: View of Asia, providing high detail of disease hotspots. Local disease events occur in clusters, separated by time gaps (1). Possible H5N1 route to Indonesia from mainland (2). Three strains are concentrated in Southeast Asia (3).

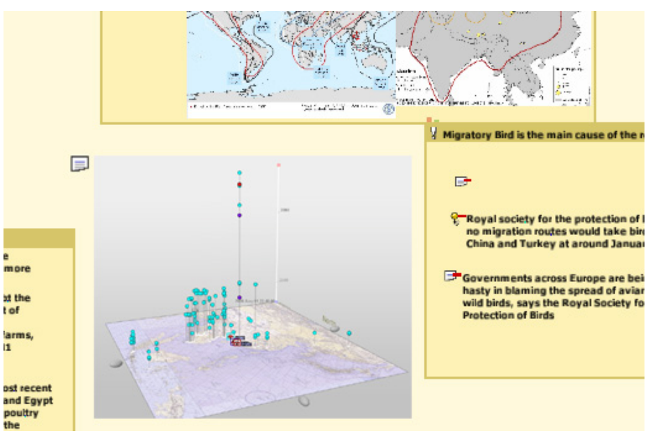


Figure 10: Linked GeoTime thumbnail in the Sandbox

3 Structuring and Cataloguing Data

As Lynn finds relevant information in TRIST and GeoTime, she integrates it with her existing hypotheses in the Sandbox. As the Sandbox does not force the analyst into a particular model or mode for the layout of their thoughts and evidence, Lynn is free to arrange information as she sees to encode meaning. For example, she places supporting details with higher level concepts, while related issues are placed next to each other. She emphasizes items more important to her by increasing their relative size. All these sense making interactions are achieved through direct manipulation using gestures. Drawing a circle around objects selects them for moving. Drawing a second circle around selected objects groups them. Items or groups can be moved and dropped on other groups. Groups can also be moved with collision detection active and will then knock other groups out of the way when the group is placed. Editor gestures insert and delete Sandbox space by animating objects out of the way or bringing them closer together.

As Lynn moves between information retrieval and analysis, a diagram of the situation and assessment begins to emerge, as shown in Figure 11 with details in Figure 12.

Figure 11: The emerging analysis diagram in the Sandbox.

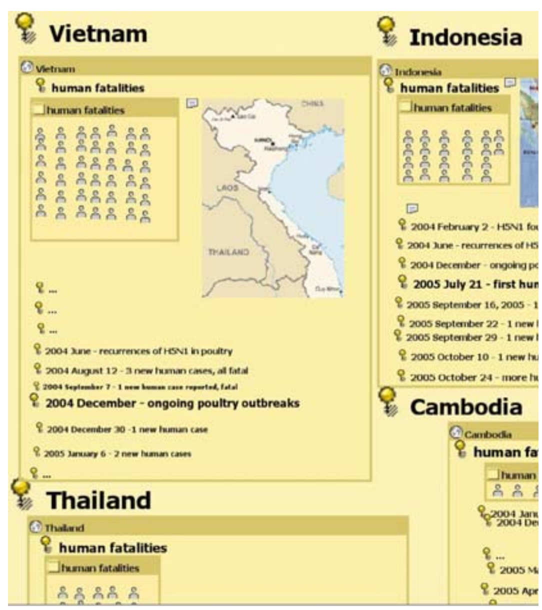


Figure 12: Analysis diagram in detail

While Lynn collects various types of information pertinent to her analysis, she notices a general hypothesis among experts that the virus might be spread by migrating birds. Lynn makes a group in the sandbox and saves all the relevant links in this subject into the group so she can come back to the topic in the future, as seen in Figure 13.

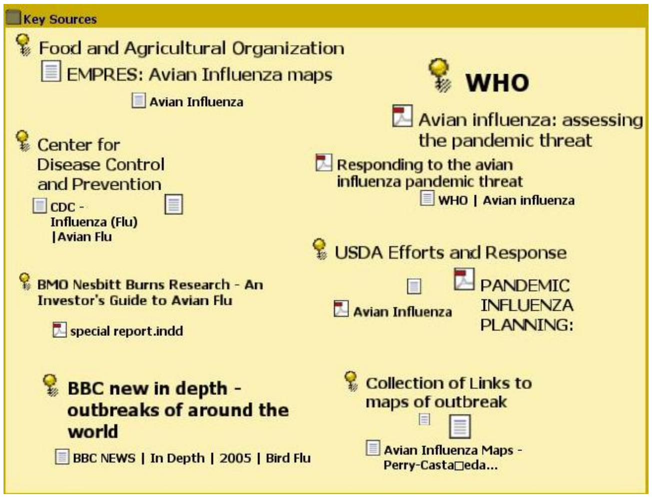


Figure 13: Important links related to a single hypotheses

3.1 Generating and Testing Hypothesis

Lynn uses assertions to keep track of the arguments for and against the proposed measures to control the propagation of the disease. Assertions have evidence gates that encode the object dragged in as supporting or refuting evidence in a single gesture, as shown in Figure 14. Assertions can also be nested to create inference networks. An at-a-glance indicator on top of each assertion provides quick assessment of the quantity and type of evidence retrieved so far.

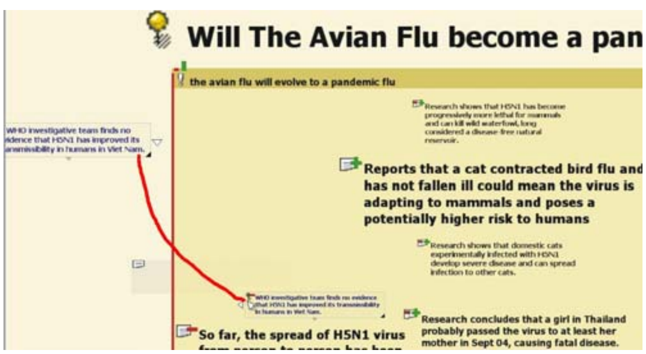
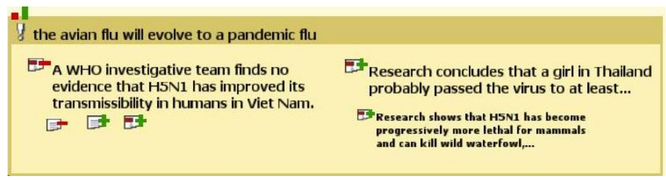


Figure 14: An Assertion with Evidence Gates. As the analyst drags evidence through the left refuting gate of an assertion, the gate flashes red, the evidence is tagged with a minus sign and the red visual indicator on top of the assertion grows. A corresponding supporting gate, on the right side, flashes green, adds a plus sign and updates the green bar of the assertion score when used.

The explicit visual thinking supported by the Sandbox helps counter natural human tendencies to only look for evidence that supports the hypothesis. These tendencies highlight evidence which sounds more plausible in the analyst’s mind or the cognitive bias. This has the negative effect of giving more strength to certain arguments based on how recently they were processed. Assertions can be combined into an analysis of competing hypotheses (ACH) framework [9]. Visual cues indicate diagnosticity of the ACH evidence. An evidence item used in many alternatives is blurred while an item uniquely supporting one alternative hypothesis has a sharp icon.

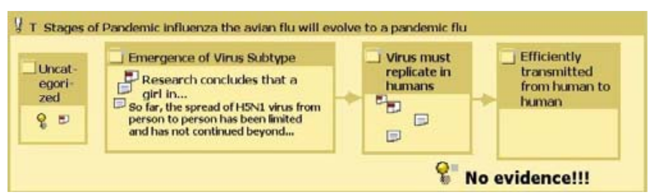
3.2 Exploring Alternative Perspectives

Analytical models that encapsulate expert knowledge and best practices can be quickly applied and un-applied to a situation to provide an alternative point-of-view. To reduce the cognitive biases associated with a particular mindset, exploring different perspectives is key to successful analysis of a scenario. Trying to fit the collected evidence to different models might reveal gaps or different interpretations of the evidence. Lynn takes advantage of her colleague's knowledge captured in a model he created during his investigation of the AIDS epidemic in Africa. The cycle of poverty and hunger applied to her assertion about the threat of the disease propagating faster once in Africa provides new insight, which suggests that impoverished people often live in close proximity to poultry, and are unwilling to kill their livestock due to their reliance on them for food and income. Another model, which structures evidence as a process diagram, highlights the fact that as long as the virus does not propagate between humans the risks are low. These steps are highlighted in Figure 15.



Step 1. Exploration workspace, in this case an assertion.

Step 2. Process Template on the Stages of Pandemic Influenza.



Step 3. Applying the Pandemic template automatically placed the evidence from the exploration workspace in this new perspective.

Figure 15: Using analytical templates for alternative perspectives.

3.3 Filling in Gaps

Zooming out to review her overall analytic progress, Lynn realizes that her assertion regarding the disease being propagated by the migrating birds is under-developed. She decides to investigate migration routes in GeoTime as a possible cause of disease transmission. Lynn augments her original GeoTime view with additional information animal migratory routes, and color codes events by outbreak type. She organizes each animal type in a layer so that the events associated with each type of animal can be easily toggled on or off. Lynn focuses her analysis on the most widespread strain of avian flu, H5N1, which was localized to the eastern hemisphere. She adjusts the visualization configuration to include human and bird cases only, as shown in Figure 16.

It becomes visually apparent that the reported migratory and wild bird cases do not frequently coincide with farmed bird or human cases. This analysis lends credibility to the hypothesis that migratory birds have not been a significant factor in the spread of the disease. She annotates and drags this view in the corresponding assertion in the Sandbox, entering through the left gate, which tags the observation as refuting evidence and recalculates the overall strength of this assertion, all in one simple gesture.

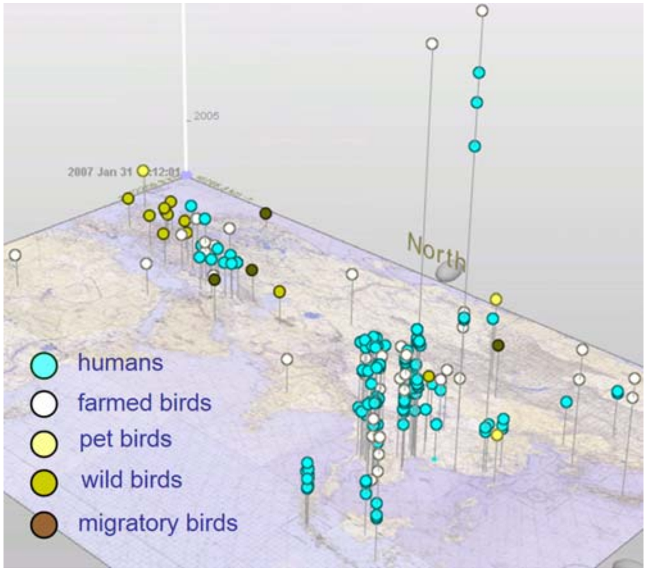


Figure 16: H5N1 outbreaks colored by outbreak type.

However, Lynn notes that wild bird testing is a relatively recent procedure, which suggests that true wild and migratory bird infections may be far greater than reported. Since it has already been established that avian flu occurs naturally in certain waterfowl, she decides to further evaluate the role of migratory birds by overlaying migration routes in the existing visualization.. She quickly finds open source information on migration routes using TRIST and uses inking functionality in GeoTime to sketch these routes in the same context as the event data, as seen in Figure 17.

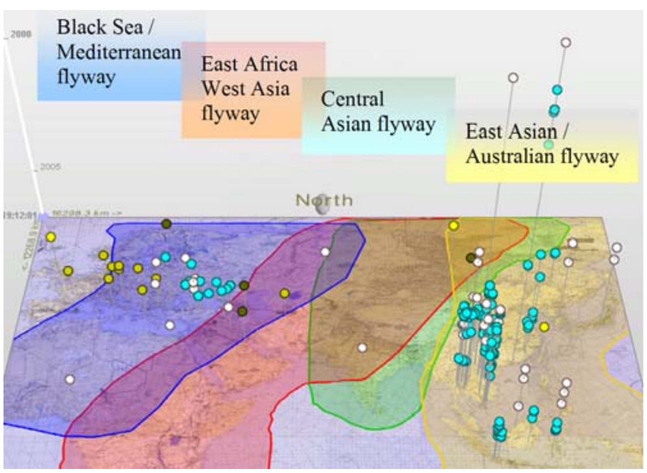


Figure 17: Bird and human cases and migration patterns of the eastern hemisphere.

The resulting view in GeoTime reveals two distinct outbreak regions, each constrained within a migration route. The hotspot events of Figure 9 are entirely contained within the bounds of the East/Asian Australian flyway. This may also provide an explanation of how H5N1 reached Indonesia, an area isolated from mainland Asia. The European and Middle Eastern outbreaks appear within the Black Sea/Mediterranean flyway, and disease could be spreading due to this migration route. The overlap of the zones display a possibility that the H5N1 could spread to different flyways, although the distribution of events, as seen in GeoTime, suggest this has not happened yet. A single bird outbreak in India is the only outlier, existing in a flyway otherwise unaffected by H5N1. Investigation of the source of the Indian outbreak is therefore of interest to Lynn.

Unlike Figure 16, Figure 17 provides strong evidence linking bird migration to the distribution of H5N1. Examining the temporal dimension, Lynn can see that the disease was restricted to the core of the migration routes, until more recently, when disease events occur in regions overlapping with other migration routes. H5N1 has just begun to enter into these previously untouched routes, and may therefore spread within the new routes in the coming season as a result of spring migration. Monitoring of wild and migratory birds along these routes in the future would be advisable. More careful examination of the East Asian/Australian flyway reveals that this route covers an area of Alaska. This provides a way for H5N1 to enter into North America, as shown in Figure 18.

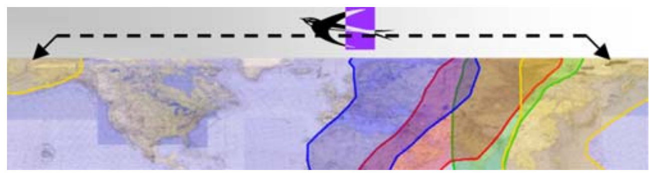


Figure 18: Migration route linking Asia with North America.

Since Alaska is socially isolated from the core North American population, and from poultry trade, an Alaskan H5N1 outbreak which precedes any other North American outbreak would be strong evidence that H5N1 is spreading via bird migration. The results of this analysis might effectively be used to persuade Alaskan authorities to conduct sentinel surveillance of wild birds to monitor for the presence of the virus. The United States Fish and Wildlife Service already have a strong presence in Alaska, and their services can be requested to conduct periodic testing of wild birds. Lynn decides to also setup a monitoring query in TRIST to keep an eye on Alaska.

At this stage of the analysis process, Lynn feels that her assessment is clear enough to invite her colleagues to review it. Since the work and the results are readily accessible in their visual form, the presentation yields productive results by affording complimentary feedback.

Lynn's final report does not require very much attention until the last minute since she knows it can quickly be generated by moving selected elements from the Sandbox to MS Word or PowerPoint, which also automatically formats the document and generates appropriate references. Lynn may also use the Sandbox itself to create an interactive final report.

4 Discussion

This case study is an attempt to investigate and validate approaches to visual analytics that support the analysis of complex, real-world scenarios. Our aim in conducting this case study is to gain a deeper understanding of the analytic process, and how our tools support this process, so that future tools may benefit from the insights generated through our study.

It is clear from the results of the case study that visual analytic tools that aim to support an analytic workflow must be flexible in order to enable the freeform representations of thoughts, hypotheses, and unstructured data. Providing the analyst with a flexible workspace enables the free-thinking necessary for effective analysis.

It is also important that these tools support uncertainty and ambiguity, as often there is no single correct answer or interpretation of evidence. Furthermore, we believe that it is beneficial for the analyst to observe and understand uncertainty within the context of hypotheses and findings, thus enabling quick interpretation in order to make reliable assessments.

Finally, we believe that visual analytic tools should enable quick and easy export to existing applications, such as MS Word and PowerPoint, in order to fit within existing analytic workflows and make the process of generating a report or presentation seamless.

5 Future Work

In the future, we plan on examining how data can be more seamlessly transferred from one analytic environment to another. A difficult problem involves transforming unstructured data accumulated in a freeform environment, such as nSpace, to a more structured environment, such as GeoTime. We are currently investigating mixed-initiative techniques for semi-automated event extraction, along with other techniques to facilitate fluid data transfer.

Future work also includes building on the best of nSpace as a unified analyst environment and as an open integration platform to implement a distributed infrastructure in which analytic components can interoperate and provide capabilities to each other as well as to applications for analysts. One example of ongoing work in this area is an effort to integrate components that will support analysts that work with information in multiple languages.

6 Conclusions

This case study illustrates how nSpace and GeoTime can be combined to create a powerful analytic setting that supports the analysis of complex, real-world scenarios. While the synergy created though the integration of these tools enabled Lynn to quickly understand the scenario, generate meaningful hypotheses, and present her results, there remain several avenues for improvement, specifically the transfer of data between different analytic environments. Nevertheless, the benefits of analytic tool integration cannot be contested, and based on the results of this case study, we believe that supporting a fluid, non-sequential workflow will lead to performance benefits for the analyst.

7 Acknowledgements

This study was supported and monitored by the Advanced Research and Development Activity (ARDA) and the National Geospatial-Intelligence Agency (NGA) under Contract Number NMA 401-02-C-0032 and NMA 401-02-C-0015. The views, opinions, and findings in this report are those of the authors and should not be construed as an official DOD position, policy, or decision, unless so designated by other official documentation. The authors wish to thank all ARDA / DTO staff for their support and encouragement.

References

[1] Center for Disease Control and Prevention, "Epi Info Features", Retrieved 20 March 2006 *from http://cdc.gov/epiinfo/*

[2] Davidson, D., "Essays on Actions and Events", Oxford: Clarendon Press, 1980.

[3] Freitag, D., M. Blume, J. Byrnes, R. Calmbach, and R. Rohwer, "A Workbench for Rapid Development of Extraction Capabilities", 2005 International Conference on Intelligence Analysis (analysis.mitre.org), McLean, VA (2005).

[4] Geroimenko, V. and C. Chen, "Visualizing the Semantic Web", Springer-Verlag, London, 2003.

[5] Green T. R. G., "Cognitive Dimensions of Notations", Proceeding of the HCI '89 Conference on People and Computers (1989), 443–460.

[6] Harabagiu, S., D. Moldovan, C. Clark, M. Bowden, A. Hickl, P. Wang, "Employing Two Question Answering Systems in TREC 2005", Proceedings of the Fourteenth Text REtrieval Conference (TREC 2005), Gaithersburg, Maryland, November 15-18, 2005.

[7] Hearst, M., "User Interfaces and Visualization", Chapter 10 in Ricardo Baeza-Yates, Berthier Ribeiro-Neto, Modern Information Retrieval, Addison-Wesley-Longman, 1999.

[8] Hendry D.G., D.J. Harper, "An Informal Information-Seeking Environment", Journal of the American Society for Information Science. 48, 11 (1999), 1036–1048.

[9] Heuer, R. J. Jr., "Psychology of Intelligence Analysis", Central Intelligence Agency: Center for the Study of Intelligence, 1999.

[10] i2 Limited., http://www.i2inc.com

[11] Institute for Human Machine Cognition, University of West Florida., Cmap Tools, *http://cmap.ihmc.us/*

[12] Jonker, D., W. Wright, D. Schroh, P. Proulx and B. Cort, "Information Triage with TRIST", In Proc. International Conference on Intelligence Analysis, 2005.

[13] Kapler, T. and W. Wright, "GeoTime Information Visualization", Information Visualization Journal, Palgrave Macmillan, 4,(2), Summer 2005, 136-146.

[14] Ladd, C., "An Overview of Public Health Preparedness Epidemiology", Houston Department of Health and Human Services, Bureau of Epidemiology, 2005.

[15] Mindjet Inc., *http://www.mindjet.com*

[16] Morse, E., M. Potts Steves and J. Scholtz, "Metrics and Methodologies for Evaluating Technologies for Intelligence Analysts", In Proc. Conference on Intelligence Analysis, 2005.

[17] Pirolli, P. and S. Card, "Information Foraging", Psychology Review Vol. 106, No. 4, 1980, (pp.643-675).

[18] Scholtz, J. et al, "Pilot Evaluation of TRIST", Internal NIST Technical Report, February, 2004.

[19] Visual Analytics Inc., http://www.visualanalytics.com

[20] World Health Organization, "Avian Influenza Timeline", Retrieved 22 February 2006 from *http://www.who.int/entity/csr/disease/avian_influenza/timeline.pdf.*

[21] World Health Organization, "Disease Outbreak News", Retrieved 22 February 2006 from *http://www.who.int/scr/don/en/*

[22] Wright, W., D. Schroh, P. Proulx, A. Skaburskis and B. Cort, "The Sandbox for Analysis – Concepts and Methods", ACM CHI 2006.