A Web-Based Task-Tracking Collaboration System for the Florida Public Hurricane Loss Model

Raul,Garcia*, Diana Machado*, Hsin-Yu Ha*, Yimin Yang*, Shu-Ching Chen*, Shahid Hamid†

*School of Computing and Information Sciences, Florida International University, Miami, FL 33199, U.S.A.
{rgarc101,dmach009,hha001,yyang010,chens}@cs.fiu.edu
†Department of Finance, College of Business, Florida International University, Miami, FL 33199, U.S.A.
hamids@fiu.edu

Abstract—The Florida Public Hurricane Loss Model (FPHLM) is a large scale, multidisciplinary project developed to assist the state of Florida with the regulation of residential insurance premiums as they relate to insured losses caused by hurricane winds. The modeling services provided to clients using the FPHLM involve physically distributed personnel with different levels of technical expertise. Bringing together such a team to collaborate on the complex task of operating the FPHLM requires a centralized system to support the effective allocation of human resources, the tracking of the different stages of the data processing, and the exchange of information among team members and between the modeler and the clients. In this paper we present a web-based collaboration system that automates the FPHLM's insurance data processing workflow.

Keywords—distributed data processing, task-tracking, workflow automation, Florida Public Hurricane Loss Model

I. INTRODUCTION

Natural disasters such as hurricanes may cause tremendous economic and human losses. In 2005, for instance, Hurricane Katrina caused an estimated damage of $156 billion [1]. In more recent years, 2011's Hurricane Irene caused nearly $16 billion in damage [2], and just last year Hurricane Sandy caused damage estimated in $50 billion [3]. Much work has been done to provide solutions for emergency management during disaster recovery [4], [5]. On the other hand, to help reduce these catastrophic losses, it is critical to be able to predict the impact of statistically likely hurricanes on the current building stock. As the only public hurricane loss projection model in the country, the FPHLM has been developed to assist the state of Florida and the residential insurance industry with the ratemaking process, and it has successfully served Floridians for almost a decade [6].

The operation of the FPHLM to provide modeling services is a highly collaborative activity that involves geographically distributed actors, from the client who requests the service to the multiple members of the modeling personnel who participate in the generation and delivery of the results. In order to facilitate a seamless collaboration among participants and better coordinate the insurance data processing workflow of the FPHLM, we have designed and implemented a web-based system that follows a three-tier software architecture pattern for easy extensibility and maintainability.

Our system automates most of the FPHLM data processing workflow by providing tailored features for task management, which encompasses the creation, assignment, tracking, configuration, and execution of modeling tasks. It also provides features for event notifications, model configuration, data analysis aided by multiple attribute frequency distribution tables, authentication and authorization management using access control tables based on fine-grained system actions and user roles, and delivery of results to clients. The integration of all these features into a coherent online collaboration system amounts to a better and faster service to industry clients and state regulators.

Our system automates most of the FPHLM data processing workflow by providing tailored features for task management, which encompasses the creation, assignment, tracking, configuration, and execution of modeling tasks. It also provides features for event notifications, model configuration, data analysis aided by multiple attribute frequency distribution tables, authentication and authorization management using access control tables based on fine-grained system actions and user roles, and delivery of results to clients. The integration of all these features into a coherent online collaboration system amounts to a better and faster service to industry clients and state regulators.

Our system automates most of the FPHLM data processing workflow by providing tailored features for task management, which encompasses the creation, assignment, tracking, configuration, and execution of modeling tasks. It also provides features for event notifications, model configuration, data analysis aided by multiple attribute frequency distribution tables, authentication and authorization management using access control tables based on fine-grained system actions and user roles, and delivery of results to clients. The integration of all these features into a coherent online collaboration system amounts to a better and faster service to industry clients and state regulators.

II. RELATED WORK

In this section we first provide a background introduction on related works concerning hurricane loss modeling and their challenges. Next, we present a literature review of general task-tracking and collaboration systems.

A. Hurricane Loss Projection

Since natural disasters like hurricanes pose a huge threat to human life and the stability of the economy, they draw considerable attention from the insurance industry and a wide range of academic research areas. Both industry and academia devote significant resources to the development and

improvement of models that can better predict the damage that these phenomena can cause in order to understand how to best direct their mitigations efforts. [7] points out that hurricane loss estimation models are among the most common disaster simulation models, which have enormous impact and involve multidisciplinary efforts from fields like meteorology, engineering, statistics, and actuarial science.

The design and implementation of a hurricane loss projection model requires the close collaboration of experts with sharp data analysis skills in order to iteratively refine the model predictions. Given the complexity of such a model, researchers usually start by decomposing the larger problem into more focused, domain-specific sub-problems [8]–[10] that once calibrated can be made to work within a larger framework. In [8], Powell et al. with the Hurricane Research Division (HRD) of the National Oceanic and Atmospheric Administration (NOAA) presents an application called Wind Analysis Distributed Application (WAnDA) to closely analyze tropical cyclone surface wind observations in real-time. This system, which was used during Hurricane Fran, proved its effectiveness in building an accurate forecast and warning cycle.

In [9], the researchers model the radial profiles of sea level pressure and winds in the presence of a hurricane. Twelve hurricanes were used to evaluate the proposed models ability to reproduce hurricanes profiles. In [10], two key modeling components in the hurricane simulation process the wind-field model and the filling model were thoroughly investigated and their ability to precisely compute hurricane wind speeds was evaluated.

Analyzing the relationship among different factors is also one of the major research topics in modeling hurricane losses [2], [11]. For instance, Wang et al. [11] analyze the relationship among the size of a hurricane, its wind speed, and its hazard levels. Murnane et al. [2] analyze the losses caused by Hurricane Ivan in 2011 and point out the exponential relationship between the increasing wind speed and the induced losses. The authors also indicate that the exposed property and population characteristics constitute a major factor in the amount of future losses.

The research works presented so far aim primarily at refining one or two domain-specific components and their operation usually requires well-trained and experienced researchers.

Contrary to the aforementioned models, the HAZUS-MH Hurricane Model Methodology proposed in [12] incorporates experts from multiple disciplines and it is made up of five components: hurricane hazard model, terrain model, wind load model, physical damage model, and loss model. This model adds a new step to the process of estimating the hurricane-induced damage and losses rather than modeling the loss based on the given wind speed and historical loss data. In [13], a storm loss prediction model is implemented based on high-resolution insurance loss records. This model can estimate losses in two different scenarios: given a spatial level of administrative districts or for a single storm event. With the specific regional loss characteristics, the proposed model aimed at replicating the losses. However, it is not very successful at projecting aggregated nationwide losses.

B. Task-tracking Collaboration System

Task tracking management systems are widely used in large-scale projects for various purposes [14]–[16]. When projects involve the collaboration of geographically distributed members, who may even belong to different organizations, it is critical to monitor individual progress and keep all the involved project members aware of the overall progress at all times. Ineffective communication among members may result in costly delays that can ultimately lead to the failure of the project.

Richardson et al. [14] propose a task-tracking management system as a type of social network platform. Its goal is to attract personal and business users to post their tasks, which are publicly available for viewing. Users participating in specific tasks can work either individually or as part of a team. The system does not provide a communication platform for users to interact in real time, but it tracks the progress of tasks from inception to completion.

Cohen et al. [15] points out the importance of raising the awareness of the status of a task and its role in enhancing the collaboration among users involved in it. They propose to simultaneously display the information in two regions: one showing a list of users with their most recently completed tasks, and another one showing all the related data with a list of performed activities.

Romano et al. [16] emphasize how a large-scale project can benefit from an efficient collaboration process and propose a software prototype called Collaborative Project Management Software (C-PMS). The authors propose a five-level hierarchical framework for collaboration; however, since the proposed framework was not fully implemented, its feasibility and performance could not be validated by a case study.

III. OVERVIEW OF THE FPHLM PROJECT

The FPHLM project is a probabilistic catastrophe model that estimates insured personal residential losses in the state of Florida caused by hurricane winds. The model is developed and maintained by a multidisciplinary team of experts from the fields of Meteorology, Structural Engineering, Statistics, Actuarial Science, and Computer Science. The project involves the collaborative efforts of over 30 researchers and students from Florida International University, Florida Institute of Technology, University of Florida, Florida State University, National Oceanic Administration Hurricane Division, University Of Miami, and AMI Risk Consultants.

At any point in time, the FPHLM has a development and a production version. Every two years the development version is submitted for certification by the Florida Commission on Hurricane Loss Projection Methodology, and upon certification, it becomes the next production version used to provide modeling services to the Florida Department of Insurance Regulation (FL OIR) and private insurance companies. Since the initial production release in 2006, the model projections have been used in over 600 occasions by the FL OIR to evaluate rate change requests from insurance companies.

The model is composed of three main components: the Wind Module, the Vulnerability Module, and the Insured Loss Module. Figure 1 shows these components and their

interactions as well as the groups involved in the development of each component.

IV. INSURANCE DATA PROCESSING MANAGEMENT SYSTEM

A. Data Processing Manager

Providing modeling services is a highly collaborative activity with each service instance involving at least five geographically separated actors. Approximately half of all project members are directly involved in data processing for modeling services on a regular basis.

Due to the number of geographically distributed personnel involved in the project, efficiently operating the software implementation of the FPHLM to provide services to clients requires a centralized system that supports asynchronous collaboration and exchange of information. Additionally, given the various levels of technical expertise of the users of the system, it is essential that the user interface provide a uniform environment that facilitates the interaction with the system.

Figure 2 illustrates the flow of information among the main actors involved in the processing of an insurance portfolio through the FPHLM. Modeling requests come from two different sources: the FL OIR and private insurance companies. Although these two types of clients are typically interested in slightly different levels of details about the model projections and, therefore, the requested output files may vary, the processing of the data remains largely the same.

The modeling service starts with the client sending one or more data files to the Computer Science group for processing via secure e-mail or postal service using encrypted media. These data files contain information about the coverage limits and structural characteristics of residential properties in the state of Florida. Upon reception, the data is reviewed for conformity to the model's input specifications and any omissions are resolved in close collaboration with the client. Once all data-related inquiries have been addressed, the exposure file is uploaded to the FPHLM's Data Processing Manager (DPM), and essential metadata such as the client's name, model version, requested output, and any special instructions are recorded in the system. At this moment, the dataset is in the new stage and the DPM notifies all system administrators of the event via e-mail.

A dataset being processed through the FPHLM may be in one of six stages at any given time: new, assigned, processed, verified, approved, and sent. These stages model the different processing phases a dataset undergoes before the client receives the results. Figure 3 shows an instance of a dataset that has traversed all stages of the processing and Figure 4 shows a global view of all the tasks in the system, which the user can filter and sort to easily obtain a snapshot of relevant tasks. These stages provide a quick progress overview allowing supervisors and team leaders to allocate human resources to balance workloads and to provide more accurate turnaround time estimates to clients. As shown in Figure 3, the system automatically keeps a log of the changes in status for each dataset, recording the user and timestamp of each change. As the processing unfolds, a dataset can revert to a previous state if a problem is identified, in which case the system requires a description of the problem and notifies the affected users.
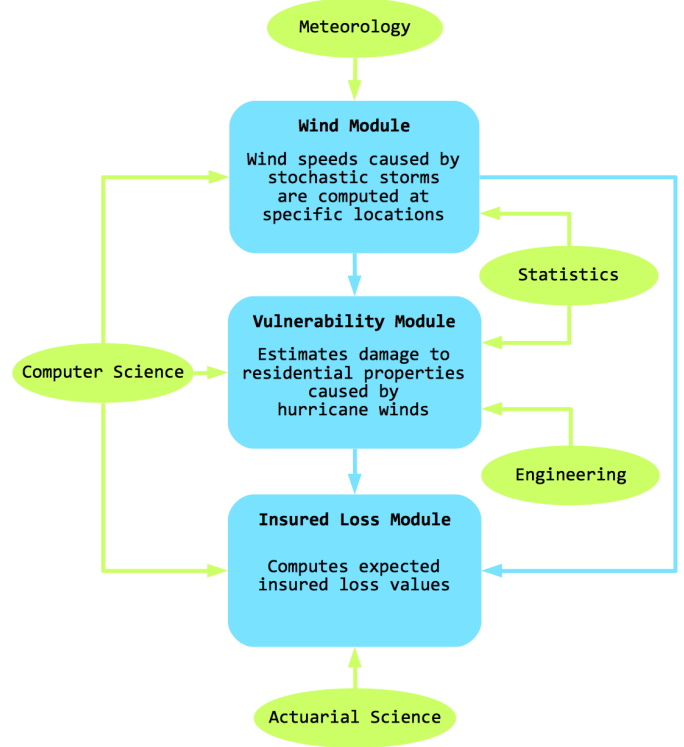

Fig. 1: Main Components of the FPHLM

Each data processing task is assigned to a team member for processing and to a second team member for verification. To facilitate the assignment process, the system displays a list of new datasets alongside a selectable list of active modeling personnel grouped by roles. Both the processor and the verifier receive automatic e-mail notifications of their new tasks, which they can review, together will all other pending tasks, in a personalized section of the system after login.

Using a large set of complex SQL functions, the insurance portfolio is preprocessed to resolve data inconsistencies, correct flawed values, assign missing critical values, and remove invalid entries. During this phase, street-level addresses are geocoded to obtain the coordinates required for the next step of the processing.

B. Wind Speed Correction Job Manager

Figure 5 illustrates the interface of the Wind Speed Correction (WSC) Job Manager. Similarly to its parent data processing manager, the WSC job manager also reports the different stages of a job: queued, running, exception, or done. These stages provide users with finer-grain status awareness compared with the six global stages of the data processing task.

The processor sets up the parameters of one or more WSC tasks using an interface similar to that of Figure 6, and adds them to a queue as a job. The system schedules jobs from this queue to be run in a cluster of enterprise-class servers and updates the jobs status accordingly. Automatic notifications are
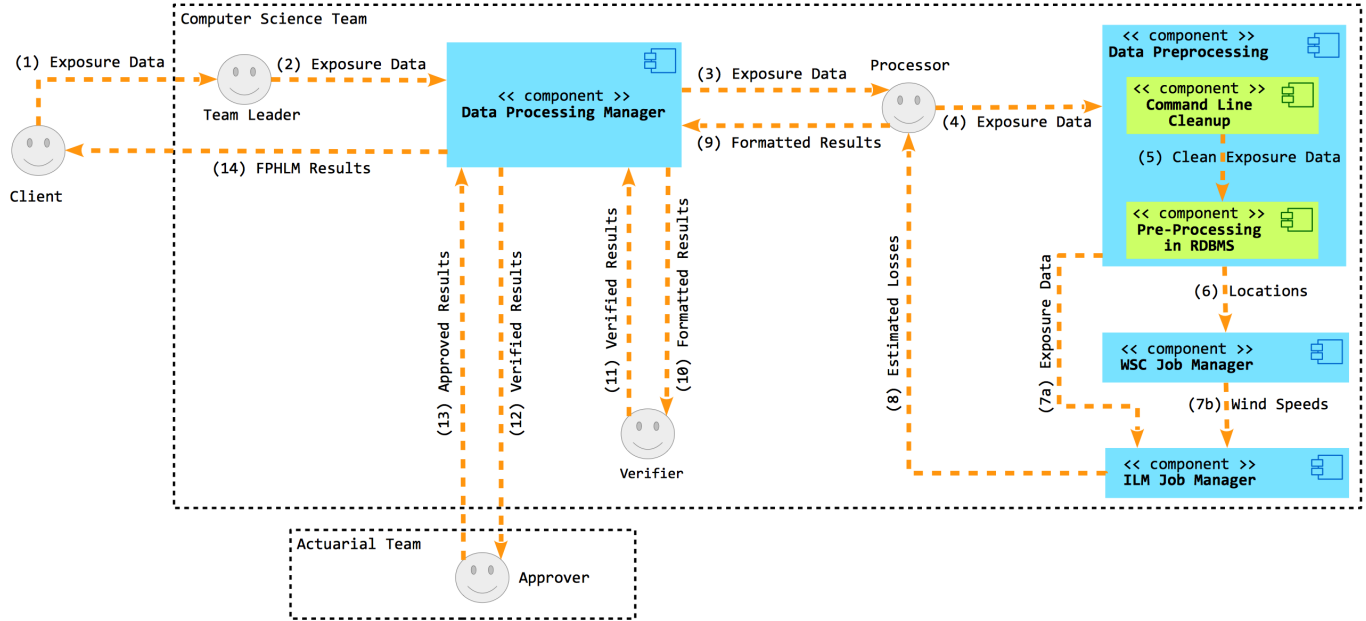
Fig. 2: Exposure Data Processing Information Flow

sent to the user who created the job in the event of an exception or upon successful completion of the job.

The output of the WSC module consists of wind speeds at different heights at each policy location for many different storm events; typically tens of thousands of stochastic events that span almost 60 thousand years of simulation. These wind speeds, together with the exposure and building characteristics of the residential properties, as well as the engineering vulnerability matrices, constitute the input of the Insurance Loss Module (ILM).

C. Insurance Loss Module Job Manager

Following a user interface design and job stages consistent with those used in the WSC Job Manager, our system also incorporates an ILM Job Manager that similarly allows users to set up the parameters, as shown in Figure 6, and schedule and track jobs using the ILM module. Just like its WSC counterpart, the ILM Job Manager automatically updates the status of the jobs and sends notifications of relevant events to all involved users.

The ILM module estimates different types of insured losses and outputs annual average losses per policy and per-event totals that are automatically formatted according to specifications.

D. Processing Verification and Results Approval

Once the processing is completed, the verifier is automatically notified, and both the processing and the results are checked for correctness. When the verifier has ensured there are no processing errors, she advances the state of the task to the next stage, which triggers the system to send a notification to the approver via e-mail.

The approver's role is to analyze the results seeking potential actuarial inconsistencies that may indicate a problem in the selection of the model parameters, modeling assumptions, or assigned values. To aid the analysis, the system provides a collection of frequency distribution tables of key attributes presented as state aggregates, or in relation with other key attributes. Once the approver is satisfied with the results, she approves them and sends a notification to the client using a built-in e-mail client that prefills all the client information and task-specific details, including a link to a personalized web page where all result files are available for download by all authorized users.

E. Results Delivery

All processing jobs belong to a client organization and users are associated with organizations. By default, all users within an organization can access result files of all the modeling requested by their organization. After a processing is completed, it is common for clients to request the results to be sent to additional users. With this setup, making results available to additional users within an organization is as simple as adding a new user account and associating it with the corresponding organization.

V. SYSTEM DESIGN AND IMPLEMENTATION

Our system manages the production usage of an active and long-term research project. Consequently, both the application domain requirements and the solution domain technologies change frequently. Requirements evolve as new research findings, data, and regulations are incorporated into the underlying model while the solution domain changes as new implementation technologies become available. To keep up with this changing environment, our system needs to be flexible and
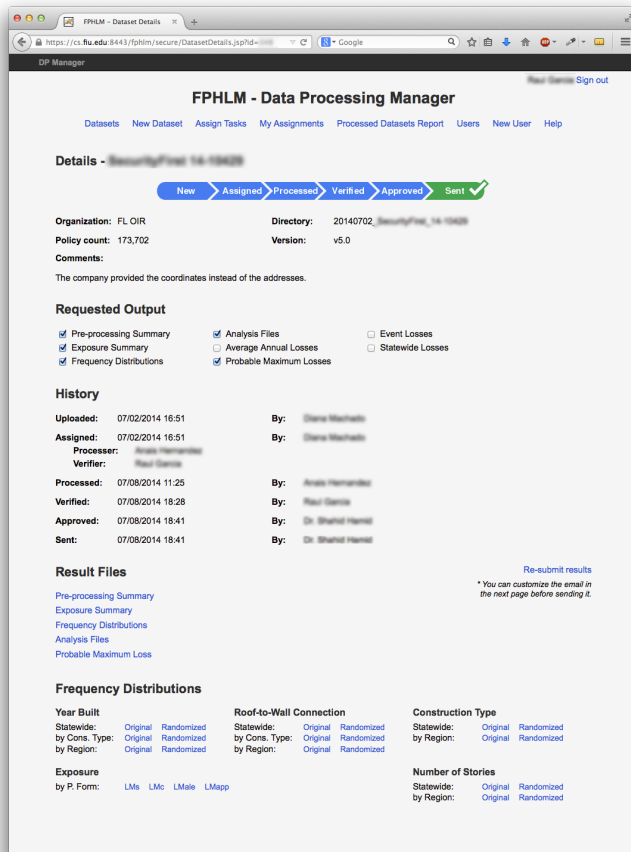
Fig. 3: Details of a Data Processing Task

Fig. 4: Datasets Overview

Fig. 5: WSC Tasks Overview

extensible without compromising stability and robustness. We chose to design our system following a three-tier software architecture pattern, which supports the notion of independent and easily maintainable tiers [17]. The popularity of the three-tier architectural pattern provides the added benefit of smoother continuity as personnel leave and join the team that develops and maintains the system. Figure 7 provides a high-level overview of the system architecture.

The presentation tier, the topmost level of the application, encapsulates the common boundary objects through which users interact with the system. In our system, these objects are standard, cross-browser controls within web forms generated by the servlet container. Figures 3-6 show examples of the user interface of the system. In this layer we employ current best practices such as separation of content and styles through the use of Cascading Style Sheets (CSS) and code reuse through the use of linked JavaScript files containing the methods to perform client-side functions such as table sorting and dynamic styling.

The logic tier of our application is made up of a web server running Apache Tomcat and a cluster of compute nodes running Linux. This middle tier provides all the functionality of the system, including task management (creation, assignment, tracking, configuration, and execution), event notifications, authentication and authorization management, data analysis,
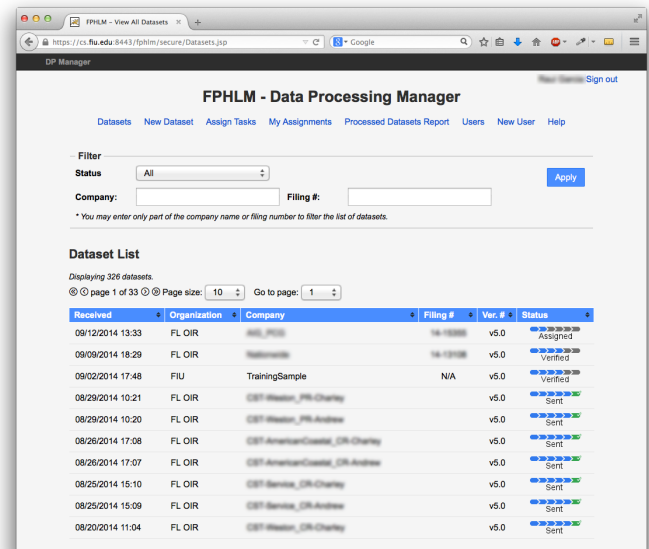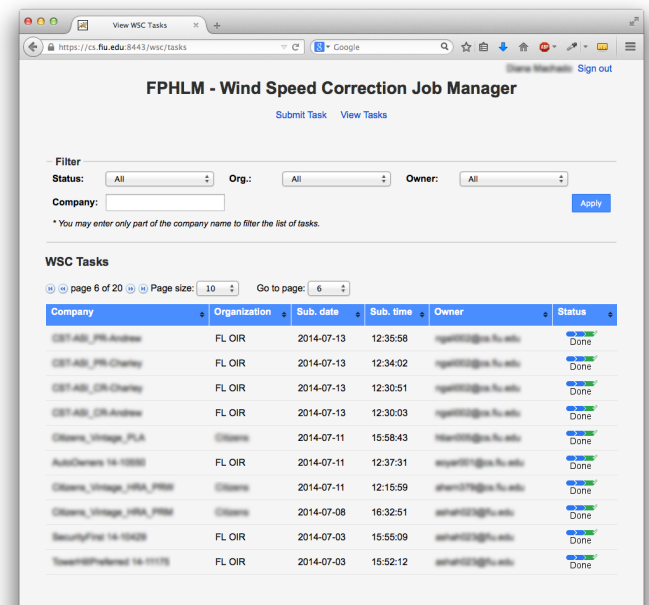
and results delivery.

Most of the application logic is written in Java while the subsystems in charge of managing the execution of the jobs are written in shell scripting languages. We developed most of the system on top of the spring framework in order to facilitate the development and maintenance of the system and reduce the effort required for testing.

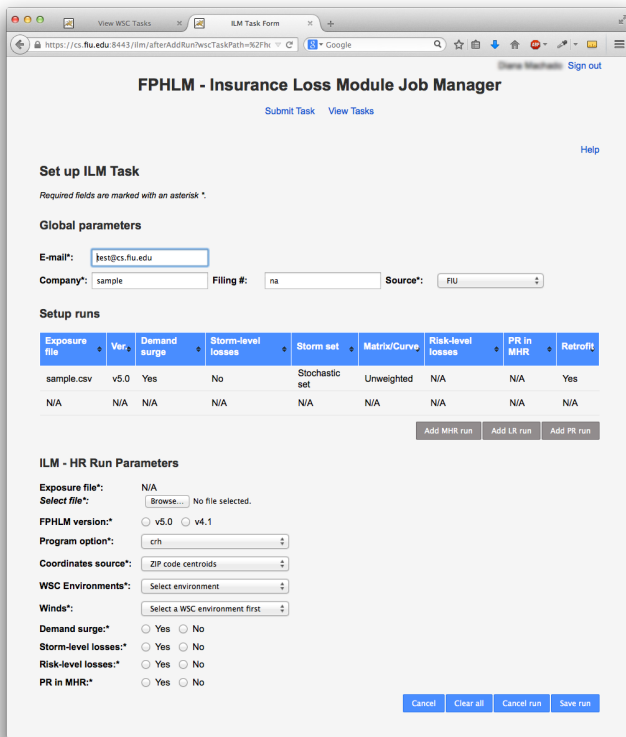The data tier of our system is composed of two PostgreSQL

Fig. 6: ILM Task Setup

Fig. 7: High-level Overview of the System Organization

coastal areas, the Florida Keys, the wind-borne debris region, and the high-velocity hurricane region.

In future versions, our system will also include a feature to overlay the estimated losses onto a Florida map, color-coded by ZIP Code or by county. The aggregated losses for all ZIP Codes or counties will be grouped into classes and will be assigned colors corresponding to different loss ranges. This type of map is helpful for evaluating the impact in losses of changing key policy attributes across multiple model runs of the same or similar datasets. Regional color-coded loss maps also provide useful insight into the insured loss distribution of a given dataset.

The current system allows the configuration of all the parameters of the components of the FPHLM; however, several of these configuration parameters remain constant across several runs of the model. In a future version, our system will provide predefined templates with default parameters for specific types of model runs. These templates will have associated authorization rules that will determine which users will be allowed to override the default value of each parameter. In addition to further easing the operation of the system, this template feature will reduce the likelihood of human errors during a data processing task.

As shown in Figure 7, our system currently stores a set of formatted result files for each data processing task. We plan on adding to the system the capability of automatically importing the raw modeling results into the RDBMS, and then create formatted result files on-demand as clients or model personnel request them using the systems web interface. This capability will greatly diversify the number of views clients have of the modeling results while demanding no additional effort from the modeling personnel.

Lastly, a feature that will also be incorporated in a future version of the system is the automatic generation of reports. Salient examples of reports include the frequency of model

RDBMS and several file servers. One of the RDBMS manages system metadata such as job states, data paths, client organizations, users, roles, and modeling output specifications. The second RDBMS houses hundreds of databases, one for each insurance dataset that is processed through the FPHLM. Each database contains several processing stages of the exposure data along with auxiliary information and scripts used during the preprocessing of the data. Since data requests for the same database come from several users over short periods of time, a database connection pool is used to boost the responsiveness of the system.

Figure 8 shows a succinct Entity-Relationship diagram illustrating the conceptual design of the systems metadata database. As captured by the Sets relationship set, the system logs all status changes including rollbacks a dataset undergoes throughout its processing. One salient feature of the system that cannot be captured in the ER diagram is the security check that requires a user to belong to the same organization the dataset belongs to in order to access the modeling results for that dataset.

VI. FUTURE WORK

Although the current system provides some support for data analysis in the form of frequency distribution tables, it does not provide graphical tools to aid the analysis process. We plan on developing a feature to automatically generate a state map showing the location of each policy in a dataset. This type of map will allow users to quickly identify regional patterns in the form of clusters of policies in areas of interest such as
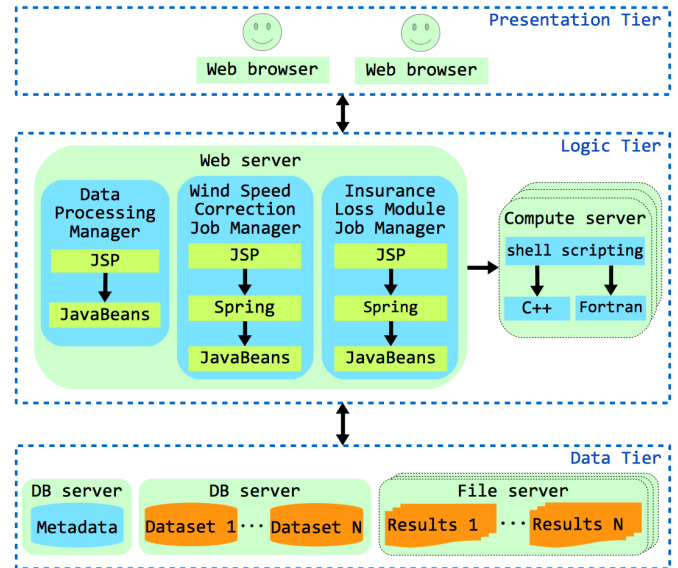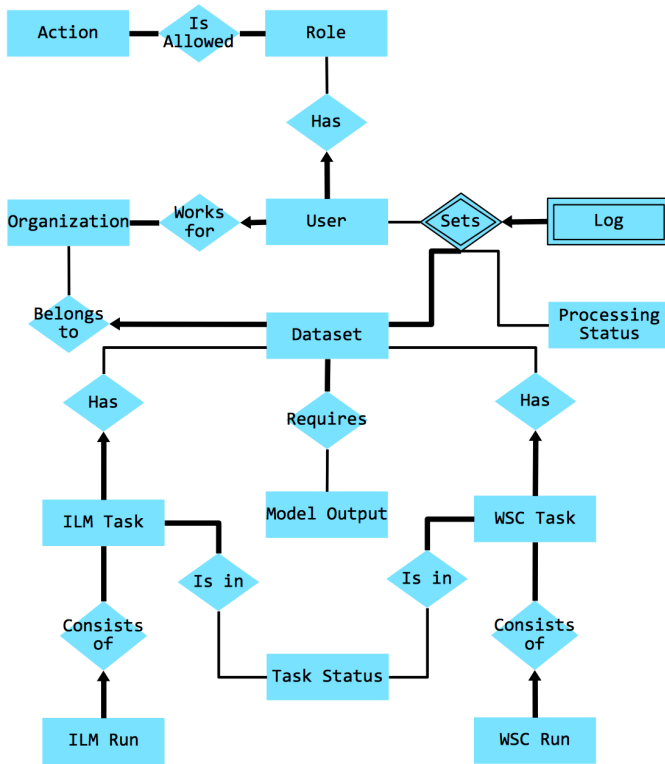
Fig. 8: System's Metadata Database ER Diagram

use, aggregated measures of insurance data processed through the FPHLM, and various system performance statistics. These reports will provide essential data to researchers developing the next version of the model, will provide better feedback to the state, and will aid the personnel in charge of the model operation in improving the performance of the models computing infrastructure. Similarly to the model results, these reports will be generated on-demand and made available from the systems web interface.

VII. Conclusion

The operation of the FPHLM is a complex activity that involves multiple geographically distributed actors with dissimilar technical skill levels, expectations, and needs. Therefore, the effective collaboration of all the members is vital for providing accurate and timely modeling services.

Our web-based, task-tracking collaboration system improves the status awareness and coordination capabilities of all the modeling personnel by providing relevant event notifications and progress information at different levels of granularity. Our system also provides support for the configuration and execution of modeling tasks as well as the analysis and delivery of their results. This support is provided using a flexible and extensible software architecture pattern that easily accommodates the changing requirements of an active research project.

To date, our system has proven extremely effective in increasing the efficiency and accuracy of the modeling tasks

of the FPHLM by roughly halving the time required to process an insurance dataset and significantly reducing the likelihood of human errors. Our system has a high potential for improvements in terms of data visualization, automatic report generation, and the further simplification of configuration tasks with the introduction of templates.

References

[1] Mark L Burton and Michael J Hicks, "Hurricane katrina: Preliminary estimates of commercial and public sector damages," *Marshall University: Center for Business and Economic Research*, 2005.

[2] RJ Murnane and JB Elsner, "Maximum wind speeds and us hurricane losses," *Geophysical Research Letters*, vol. 39, no. 16, 2012.

[3] Eric S Blake, Todd B Kimberlain, Robert J Berg, JP Cangialosi, and John L Beven II, "Tropical cyclone report: Hurricane sandy," *National Hurricane Center*, vol. 12, 2013.

[4] Yimin Yang, Wenting Lu, Jesse Domack, Tao Li, Shu-Ching Chen, Steven Luis, and Jainendra K Navlakha, "Madis: A multimedia-aided disaster information integration system for emergency management.," in *CollaborateCom*, 2012, pp. 233–241.

[5] Steven Luis, Fausto C Fleites, Yimin Yang, Hsin-Yu Ha, and Shu-Ching Chen, "A visual analytics multimedia mobile system for emergency response," in *Multimedia (ISM), 2011 IEEE International Symposium on*. IEEE, 2011, pp. 337–338.

[6] Shu-Ching Chen, Min Chen, Na Zhao, Shahid Hamid, Kasturi Chatterjee, and Michael Armella, "Florida public hurricane loss model: Research in multi-disciplinary system integration assisting government policy making," *Government Information Quarterly*, vol. 26, no. 2, pp. 285–294, 2009.

[7] Charles C Watson Jr and Mark E Johnson, "Hurricane loss estimation models: Opportunities for improving the state of the art," *Bulletin of the American Meteorological Society*, vol. 85, no. 11, pp. 1713–1726, 2004.

[8] Mark D Powell, Sam H Houston, Luis R Amat, and Nirva Morisseau-Leroy, "The hrd real-time hurricane wind analysis system," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 77, pp. 53–64, 1998.

[9] Greg J Holland, "An analytic model of the wind and pressure profiles in hurricanes," *Monthly weather review*, vol. 108, no. 8, pp. 1212–1218, 1980.

[10] Peter J Vickery and Lawrence A Twisdale, "Wind-field and filling models for hurricane wind-speed predictions," *Journal of Structural Engineering*, vol. 121, no. 11, pp. 1700–1709, 1995.

[11] Yue Wang and David V Rosowsky, "Joint distribution model for prediction of hurricane wind speed and size," *Structural Safety*, vol. 35, pp. 40–51, 2012.

[12] Peter J Vickery, Jason Lin, Peter F Skerlj, Lawrence A Twisdale Jr, and Kevin Huang, "Hazus-mh hurricane model methodology. i: hurricane hazard, terrain, and wind load modeling," *Natural Hazards Review*, vol. 7, no. 2, pp. 82–93, 2006.

[13] MG Donat, T Pardowitz, GC Leckebusch, U Ulbrich, and O Burghoff, "High-resolution refinement of a storm loss model and estimation of return periods of loss-intensive storms over germany," *Natural Hazards and Earth System Science*, vol. 11, no. 10, pp. 2821–2833, 2011.

[14] Richard A Richardson and Pedro DeAbreu, "Task completion tracking and management system," Sept. 21 2012, US Patent App. 13/624,005.

[15] Andrew L Cohen, Bob Stachel, and Steve Foley, "Method and software for supporting improved awareness of and collaboration among users involved in a task," Jan. 14 2003, US Patent 6,507,845.

[16] Nicholas C Romano, Fang Chen, and Jay F Nunamaker Jr, "Collaborative project management software," in *System Sciences, 2002. HICSS. Proceedings of the 35th Annual Hawaii International Conference on*. IEEE, 2002, pp. 233–242.

[17] Bernd Bruegge and Allen H. Dutoit, "System design: Decomposing the system," 2010.