PrIMe: System Approach to Predictive Modeling of Combustion

M. Frenklach, Z. Djurisic, M. Gutkin, D. Yeates, and X. You

Department of Mechanical Engineering
University of California, Berkeley, CA 94720-1740, USA

1 Introduction

Optimizing combustion efficiency and understanding the mechanisms that prevent full energy utilization of fuels relies on detailed knowledge of the underlying physics and chemistry. These systems are generally complex enough that models have been used to explore the effect of different feed and reactor conditions and have been successful in optimizing fuel mixtures and combustor performance. However, the models are extremely complex and often controversial. The data, which parameterize the models and are compared to model predictions, are themselves complex and often open to interpretation. Further, they are developed by multiple labs using different technologies. To keep track of models, parameters, and data in an integrated framework has proven a necessity in the field of Combustion. The initiative we call PrIMe is designed to fill this need. In its scientific content, PrIMe is a *system approach* aimed at establishing the infrastructure, both scientific and cyberinfrastructure (CI), in support of developing *predictive* models of combustion. The present article provides a brief description of the current infrastructure, focusing on new additions, primarily those to the PrIMe Data Model and PrIMe Workflow Application.

2 Major Components of PrIMe

PrIMe, which stands for Process Informatics Model (http://primekinetics.org), is a data-centric approach that relies on three major activities: collection and organization of scientific data; availability of scientific tools for effective data analysis and processing; and engagement of the entire scientific community in the data collection, curation, and analysis. The proper infrastructure must enable a new form of scientific method by considering the entire content of information available, assessing and assuring mutual scientific consistency of the data, rigorously assessing data uncertainty, identifying problems with the available data, evaluating model predictability, suggesting new experimental and theoretical work with the highest possible impact, reaching community consensus, and merging the assembled data into new knowledge and predictive models.

To support the stated objectives, the PrIMe infrastructure is composed of the following major components: Team Collaboration Portal, Data Model, Data Warehouse, Data Management software, and Workflow application. These components are described below.

Correspondence to: myf@me.berkeley.edu

3 PrIMe Portal

PrIMe, by design, provides a virtual-collaboration infrastructure. The membership in PrIMe is open to all. To facilitate curation of the data, the PrIMe team members are organized into Work Groups, each focusing on a particular aspect or part of the data. For instance, there is a work group responsible for curation of rates for oxidation reactions of C_1H_x hydrocarbons. The members of this Group review the submitted public data of the Warehouse and collectively reach consensus on the "best current recommendations" for the individual reaction rates. It is these recommendations that are then to be used by analysis and modeling.

The communication among the group members is supported by the PrIMe portal (http://primekinetics.org), built on the Drupal platform (http://drupal.org), and housed at a commercial site (http://www.godaddy.com). There are currently over 200 members registered, from about 15 countries, forming over 20 Work Groups. These Groups cover species thermodynamics, reactions of aliphatic and aromatic hydrocarbons, formation of nitrogen oxides, soot, and experimental data in various flame configurations. There is also an Industrial Work Group. Each such Group is led by an expert in the respective area, and the Group leader has administrative privileges to control the content of the Portal Group space. In other words, the operation of individual groups is completely autonomous, each setting its own rules and protocols. Leaders of all Work Groups form the PrIMe management team.

4 PrIMe Data Warehouse

The PrIMe Data Warehouse is a living collection of data pertinent to the field of combustion and to the development of predictive combustion models. The data collected include bibliography, chemical elements, chemical species and their thermodynamics (and transport) properties, reactions, reaction rate coefficients, reaction models, as well as experimental observations. The emphasis in organizing the data is on documenting data uncertainty and data provenance.

One of the distinctive features is that the Data Warehouse is organized not by the source or origin of the data, typical of most other activities, but by merging the different sources together according to the scientific meaning of the data (while tracking, of course, the origin and other metadata). A direct, read-only access to this data collection is available to everybody. The submission of new data or updating existing data records is available only to PrIMe members, in order to keep the data provenance. The present collection contains over 100,000 XML files, installed and managed on a WebDAV compliant, professionally (24/7) maintained server.

5 PrIMe Data Model

To organize and curate the data in the PrIMe Data Warehouse, we have developed all the needed XML standards and schemas: the PrIMe Data Model.

PrIMe Data Warehouse consists, conceptually, of two parts: *Depository* and *Library*. The Depository is a repository of data provided by the community. The only requirement for the acceptance of the data into the Depository is the data-record completeness, which is checked electronically with the use of data schemas. The Library is a set of data evaluated by the PrIMe Work Groups. A set of primekinetics codes is designed to assist in both initial submission and evaluation of the data.

One of the principal rules of the PrIMe Data Model is "single record for every entry—no data duplication". There are several aspects to this requirement. First, there is no duplication of the same property values. Imagine if the enthalpy of formation of a species (like OH) is allowed to be recorded in numerous records. In such a case, when this enthalpy value is modified in a new study, one needs to update all the records where this number is stored, which can lead to omissions and hence inconsistency in future use of the data. PrIMe Data Model stores every entity value once and only once, with all other listings referring to the original record. In this way, updating the value requires updating just that one single record. Second, PrIMe Data Model distinguishes between primary and

derived properties. For instance, PrIMe species records do not contain species molecular weight—it is calculated on the fly from atomic masses when needed. Thus, as a general rule, PrIMe Data Model strives to record primary properties and delegate evaluation of derived properties to data management software. Finally, the above considerations argue for a single (virtual) Depository overall.

The PrIMe data management software is designed to capture and store the metadata associated with every data submission and data action (like who, when, etc). No PrIMe Data Warehouse record (file) will be erased or modified. Instead, when changed, a new record will be created and the old one moved to data attic.

Every PrIMe data entity is assigned a unique index, primeID. This was necessitated by the fact that even for chemical species the suggested "unique identifier" (like CAS numbers and InChI strings) turned out to be not fully unique for some species, and for some they are still not developed. The primeID is a string composed of one to three lowercase letters sequence, unique to a data record type ("s" for species, "rk" for reaction rate coefficient), followed by an eight digit sequence. For example, species data record for argon atom has primeID s00000049.

The additions to the PrIMe Data Model are the Data Attributes, Instrumental Model, Optimization variables, and Surrogate Models. The first of them, Data Attributes, is intended to represent a layer of abstraction from and when needed aggregation of experimental data. Data Attributes are designed with specific intention of extracting from experimental observations a feature (target), which when compared to a modeling result provides a benchmark to the model's predictive ability under a specific set of conditions. Specifying experimentally observed property of state and a type of the feature of interest—peak, dip, inflection, initial or final value—it is possible to infer Data Attribute properties such as time to ignition, either from a pressure, temperature or species trace, time or temperature of reaction onset in reactive non-catalytic removal of pollutants, or simply a value of the observed property at a point of interest, inferred from correlation of discrete data points. Definitions of the targeted features explicitly specify the numerical methods, enabling the researchers to arrive, within the limits of the numerical methods precision, at the same evaluation of the model performance.

The PrIMe Instrumental Model allows storage of the experimental data in any form, up to raw signals, with uncertainty. The Instrumental Model captures and archives the experimenter's protocols and procedures he or she used in interpreting the raw experimental data while allowing the user to view the data in a user-demanded format; a principle departure from the current approach to publishing data. The Instrumental Model also enables the automatic identification of characteristic features of the experimental traces (e.g., Data Attributes) to be used for verification, validation, and uncertainty quantification of PrIMe-built models. As an example, consider a set of formaldehyde oxidation shock-tube experiments where the transmitted laser light intensity was measured by a photomultiplier tube and the produced voltages were recorded by a computer. The corresponding Instrumental Model ─ an XML document that codes a complex expression ─ transforms these raw data into CO concentration upon the user's request.

The Optimization Variables and Surrogate Models are new PrIMe Data Model components that enable us to interface the PrIMe Data Attributes with computational tools, which can be applied to perform consistency tests, model optimization, and propagation of data uncertainty. The integration of data with numerical codes is attained via the PrIMe Workflow application described next.

6 PrIMe Workflow Application

The PrIMe Workflow Application is cyber-based infrastructure for linking all PrIMe parts together and, most importantly, automation of the data flow. The conceptual diagram is depicted in Figure 1.

The practical implementation of the PrIMe Workflow is Web-based and is built on the Microsoft platform. The application is hosted on a dedicated server. It is being designed as a user-friendly drag-and-drop application that does not require any programming experience on the part of the user. The user can start with a prebuilt workflow or create a custom one, by selecting from a set of predefined objects. Each object represents an operation; e.g., getting an experimental PrIMe Data Attribute

(target) from the PrIMe Data Warehouse, carrying out a plug-flow calculation, or performing analysis of the generated results. A snapshot of the application is shown in Figure 2.

The infrastructure is built to enable code developers to contribute their applications by implementing a predefined interface. These applications can be uploaded via a menu-driven user interface.

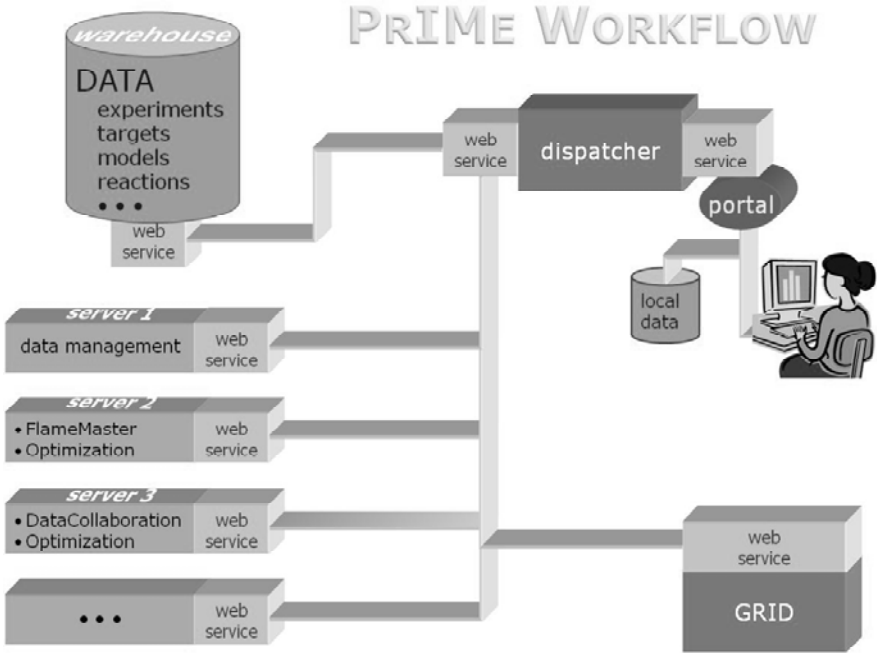

Figure 1. Conceptual diagram of PrIMe Workflow Application

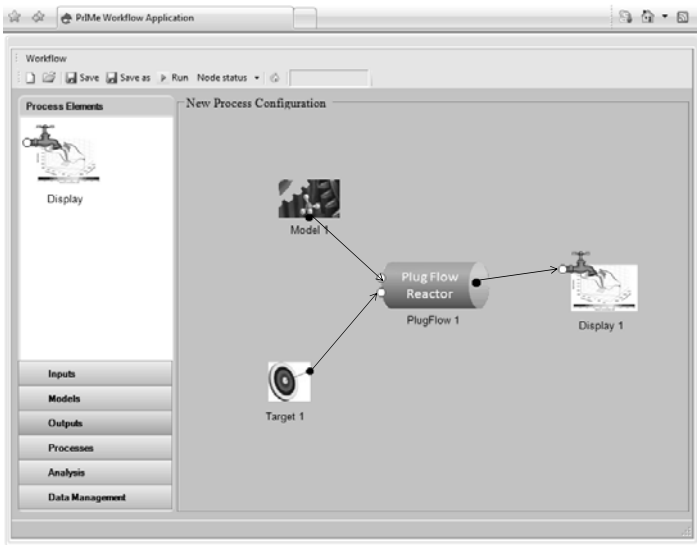

Figure 2. A snapshot of PrIMe Workflow Application

Acknowledgements. The work is supported by the NSF Chemistry Division, Cyber-enabled Chemistry, Grant No. CHE-0535542, and a joint funding by AFOSR and SERDP, Grant No. FA9550-08-1-0003, administered by Dr. Julian M. Tishkoff.