Comments for the reviewers

The authors thank the reviewers for their detailed and useful comments and suggestions. We reorganized the paper according to the suggestions and made available the software as suggested, together with sample data.
Language has been revised throughout the whole paper.

In the following, we give responses to each review in detail.

Review #1

- **Reviewer**: "Title and the use of "scholarly data": I have some doubts about the adequacy of the title to the article content. Authors use the term "scholarly data" few times across the paper, without attracting reader's attention with its meaning (at times it seems interchangeable with "scientific data" or "experimental data", e.g. p.2 l.39 or p.6 l.33) and its role in the whole system until the last section, where the use of data from publications is explicitly mentioned (as one of the features of the system only). Perhaps the title should focus more on exposing the fact of scientific data reuse or the proposed system itself."
- **Response**: The title has been modified and the terminology has been clarified throughout the whole paper. In the first version, we used "scholarly data" as a synonym of "scientific data" or "experimental data" because we consider scientific/experimental data extracted from publications. However, recognizing that this could be misleading, we changed the terminology and we now use "published scientific data" or simply "experimental data" referring to the actual data, while we use "scholarly data" only to refer to publication-related data. These terms have been clarified in the introduction.

- **Reviewer**: "Structure and content:
    - Introduction – shorter and less details (perhaps the details of the chemical kinetics domain (p.2 l. 26 - p.3 l.7) can be skipped until section 2)
    - Scenario – could benefit from clear distinction in parts about the general research area, and about the proposed approach with explicit assumptions about its functionality (i.e. naming the requirements, which can later be generalized and referred to while discussing solutions in section 3)
    - Towards an integrated framework – Despite the reference to [44], it would be nice to remind the reader what the basic requirements were (in contrast to New Requirements discussed in section 5). Are the initial functional requirements the "use cases"?
    - Proposed architecture – description of DW and OLAP systems repeats the text from section 3
    - Concluding remarks (or other section) – It would be advantageous for these considerations to elaborate a bit more on the envisioned role for the domain-specific ontology and human actors (domain experts, operators, manual actions) in the whole system – which steps cannot be automatized?"
- **Response**: The suggestions have been integrated.
    - The introduction is now shortened, and domain details have been omitted until Section 2.
    - it describes only the domain and the model development process. The emerging use cases are described starting from Subsection 4.1
    - The basic requirements were, indeed, the "use cases", and the terminology has now been clarified throughout the text to avoid ambiguities, better clarifying also the role of the previous work in the introduction:
      "Starting from a description of the most important *use cases*, which emerged also from previous works [X,X], this paper describes the development of a *new prototype* which

goes in the direction of supporting them, the *new requirements* emerged from the prototype development and an *architecture* that, taking also into account the new requirements, outlines the future development of the proposed framework."
- Repetitions from the proposed architecture have been removed, and that section has been extended to better explain the envisioned framework, also in relation to the already developed prototype.
- The concluding remarks have been extended, including also the role of the domain experts in the envisioned system.

- **Reviewer**: "Other improvements
  - The information about the existence (yes or no?) of a domain-specific ontology is inconsistent (p.12 l.36, p. 16 l.33, p. 19 l.30).
  - Figure 2 is not clear and doesn't simplify understanding the whole process: use of nodes, arrows and additional text around the nodes inconsistent (physical objects, actions, or goals) and confusing.
  - e.g.1. The central node (Exp. Data) is confusing – what does it stand for and where does "Compare' fit in the sequence?
  - e.g.2. "Analysis tools (… uncertainty quantification)" node lies on alternative path to "Model reduction" while the textual description (p.4 l.39-42) says that uncertainty is a diverging point ("if the model shows … uncertainties, relevant pathways can be identified by analysis tools ...")"
- **Response**:
  - The role of the domain-specific ontology has been clarified throughout the whole paper, as its existence at present: since a suitable open ontology does not exists for the domain at present, we developed a small and "toy" ontology just for our prototype. A future goal is to substitute it with a more complex ontology that, for example, could be extracted using automatic techniques.
    In particular, at the end of Subsection 4.2, we now better explain that we have developed a prototypal and simple ontology for our prototype, but that:
    "The ontological information developed for the current prototype could be substituted by a more complete and complex domain ontology based on a more flexible format. However, such a complete and open ontology does not exist for the domain at present and should be generated…"
    And in the paragraph "domain ontologies" from Subsection 6.1:
    "As an available domain ontology does not exists at present, semi-automated generation techniques [X] or data mining techniques to automatically generate ontological relationships based on existing data [X, X] could be used for its creation. A simple domain ontology based on human-readable files has been defined and implemented for the developed prototype, as discussed in 4.2."
  - The figure has been heavily modified following the suggestions. In particular, it has been simplified, the notation has been modified to be more consistent (data and activities use a different notation), the "target area" of our paper (validation) has been highlighted, and the figure is now consistent with the text. Moreover, the caption and the text have been modified and extended to better explain the diagram.
    Concerning model reduction, this is the last step before a detailed kinetic model is used for applications such as complex fluid dynamic simulations. The kinetic model developed and refined according to the procedure presented in the loop is, indeed, not manageable as is due to its size (number of chemical species and reactions). Model reduction, provided a desired degree of agreement with the original detailed model, allows to obtain a smaller model that is suitable for applications. Model reduction is therefore a very final step, or in other words lies outside the loop where the objective is the development of an high fidelity detailed kinetic model.

- **Reviewer** : “Language – some examples:
    - verbs in 3rd person (“leadS to” p.1 l.27, “requireS” p.7 l.2, “comeS from” p.7 l.7, “involveS” p.8 l.4, “concernS” p.10 l.21
    - typos and spelling (OpenSMOKE++ and alternatives, ReSpecTh alternatives, “Djangoi” p.12 l.24, “stred” p.16 l.23)
    - use of definite / indefinite articles
    - other: unnecessary symbols for temperature and pressure p.4 l.12, “industry industry” p.4 l.40, repeated “highlighted in the literature” p.1 l.39 and l.41, “one [variable functions]” p.4 l.32, “increasingly availability” p.6 l.32, “is not be limited” p.6 l.36, “indipendent” p.6 l.37, “as a meanS to” p.6 l.39, “it doesn’t exist” -> “there” p.10 l.3 and l.4, “a changes” p.10 l.43. “since from” p.12 l.11, “impacts on” p.16 l.19, “an automatically and manually querying” p.17 l.40,

  Some formal issues:
    - Recommended capitalisation (title, headings) missing
    - In-text citations and reference format other than suggested in the guidelines, citation starts with [32] which seems strange
    - multiple citations of some papers for the same context, e.g. [6] or [17]”
- **Response**: We revised language throughout the whole paper, taking into account all the suggestions. We changed the reference format to start with [1] and we removed unnecessary citations to multiple papers in the same context.
  However, we did not find a recommended capitalization.

Review #2

Reasons to reject:

- **Reviewer**: “I would encourage the authors to clarify the main contribution. The requirements and the architecture do not qualify as good enough contribution, worth a journal paper as 1) they have been already presented already at SAVE-SD 2018 and 2) the design of the architecture does not mean it will work in practice. I would encourage authors to go for the "first version of the system" being presented and limit the discussion of "extension to other domains and new requirements" to the discussion section, as both are not tested in practice. Right now the paper mentions them several times, but discussing the problems does not solve them...”
- **Response**: The main contributions have been clarified in the introduction:
  “Several use cases for an integrated system which leverage large amounts of published scientific data extracted from heterogeneous sources to support scientific model development are presented.
  A new prototype is developed and distributed.
  New system requirements, which emerge also from the prototype development, are identified and discussed.
  A data-based and service-based architecture for such a system is outlined and discussed, considering the new requirements and focusing in particular on the data model and the design of a set of data curation and analysis services.”
  We also highlighted the fact that the use cases, the requirements and the architecture have been revised and new requirements emerged with respect to SAVE-SD 2018, mainly thanks to the development of the new prototype which was not part of SAVE-SD.

  Moreover, following the suggestion, we clarified and distinguished the part of the framework that has been already developed (the prototype, 4.2) and the part which has been only designed (the architecture, 6).
  More specifically, we extended and revised Subsection 4.2 (prototype) to include all and only the features and the components already developed (and distributed in the source code). Now Subsection

4.2 describes the “first version of the system” and it already includes core features such as the automatic simulation of an experiment starting from some acquired data, the integration with the validation tool (Curve Matching), and front-end features such as experiments browsing.
In the new version, Section 6 (architecture) describes the overall system, which includes the features already developed in the prototype and also other features and components not developed yet, which derive from the new emerging requirements and are the goals of the ongoing research. This has been clarified at the beginning of the section 6:
“As illustrated previously in Subsection 4.2, some of the components and features have been already developed in the prototype. The following architecture includes, besides them, other components and features that must be considered as not developed yet. The development of the existing features in the prototype has proved to be crucial to refine the following architecture.”
The role of Section 6 is therefore to put the components and the features already developed and described with the prototype in the overall context of an architecture.
Of course, the solutions proposed in this section are more abstract and we can not guarantee they will work in practice since they are not developed yet, but they are the result of the experience gained during the development of the prototype and, indeed, the section has been totally re-written with respect to the preliminary solutions presented in SAVE-SD before the prototype development.
The differences between the developed prototype and the designed (not developed) architecture have been clarified throughout the whole paper.

- **Reviewer**: “A possible reason to reject would be that neither the data (or samples of data) or the software (prototype) presented in the paper are not available online. I leave it up to the editors to decide if it is a reason for rejection.For instance, when the authors describe supplementary materials in p.4, examples of such materials can be given. The same for the "experiment consists of a set of conditions, ... and output variables" - including an example would help a lot!”
- **Response**: We included the prototype and samples of the data online (GitHub). For the data imported from external repositories, we included the link.
  Moreover, we added figures 2 and 3 to better explain the conditions and the output variables when described.

- **Reviewer**: “It is not the strongest reason for rejection, but before accepting this manuscript, the paper should clarify the title and the use of terms. [1, 2] use "scholarly data" as "the data describing the publications, authors, etc.". This paper deals with the "experimental data from the chemical kinetics domain" or "scientific data". Therefore, the title "scholarly data analysis to aid ..." is misleading, as no such analysis happens at the moment (it is just mentioned as a possible source of new data). Such analysis, if already performed is non-trivial and can constitute contribution per se. For instance, on p.17, 5.3: "can be retrieved automatically from external sources". Can be? Or "is retrieved"? From which sources?”
- **Response**: Following the suggestion, we have modified the terminology and the title throughout the whole paper. In the first version, we used “scholarly data” as a synonym of “scientific data” or “experimental data” because we consider scientific/experimental data extracted from publications. However, recognizing that this could be misleading, we changed the terminology and we now use “published scientific data” or simply “experimental data” referring to the actual data, while we use “scholarly data” only to refer to publication-related data. These terms have been clarified in the introduction. Indeed, at the moment we mainly focus on experimental data extracted from publications, while the analysis of scholarly data itself is an emerging requirement.

Further comments:

- **Reviewer**: “Please give concrete examples when mentioning other domains besides "chemical kinetics": which domains, which data will be stored and managed? If you do not have anything concrete at this stage, it is better to just say "we are going to look into extending our system to work with domains A, B, C". Right now the reader might be mislead by claims like "can be easily generalized to other domains"”

- **Response:** We included more examples of other domains with similar challenges, citing relevant works and discussing briefly the similarities. In particular, we included domains such as computational biology, gene expression analysis, biomedical research, cell biology and predictive structural materials science. At the same time, we clarified throughout the paper that:
  "The framework investigated in this paper has been currently tested and optimized only for the scenario described in Section 2, but most of the design choices were taken trying to generalize its applicability"

- **Reviewer**: "The same at p.6 "The need of a continuous validation of models based on new experiments is shared among most scientific fields, and therefore the activities of acquiring, analyzing and evaluating models and experiments is certainly shared with other scientific domains" - consider giving a couple of examples to be specific."
- **Response:** See previous answer.

- **Reviewer**: "p.3, one of the contributions: "A data-based and service-based architecture for such a system is proposed and discussed, providing a data model to support data integration and the development of a set of data curation and analysis services." - please be more concrete when describing the data integration in the paper. For instance, in Fig. 1 it is not clear which steps are covered by the system, and which are automated, which are performed manually? Elaborating on this aspects and relating the authors' contribution to this general model refinement procedure would help to clarify the benefits of the presented approach."
- **Response**: We better clarified data management, both in the current prototype and in the designed architecture, respectively in sections 4.2 and 5.
  Fig. 1 has been changed. It has been simplified and the notation has been modified to be more consistent. We added also a dotted rectangle that highlights which steps of the development/validation/refinement/use cycle are the target of the proposed system. The figure has also been clarified in the text and in the caption.

- **Reviewer**: "In the related work section a good analysis of the scholarly data and chemical kinetics experimental data is presented. However, the work seems to be also very relevant to the Google Dataset Search https://toolbox.google.com/datasetsearch, Figshare and other data repositories, as well as to platforms for sharing code/experiements (MyExperiment, Gigantum, Code Ocean). I would strongly recommend relating the solution to those platforms in the Related Work section."
- **Response**: Following the suggestion, we included these relevant works and discussed how our work relate to them in Section 3.

- **Reviewer**: "Sections 4 and 5 miss the description of what exactly is done (or not done) by the described prototype and within the approach. For instance "In other cases new experiments could be directly extracted by domain experts and inserted manually" - it is not clear if this is the main source of filling the repository or only a theoretical possibility. In the former case, how often do experts add experiments manually, can they really cope with the "big data" announced earlier? Is it scalable? For the automated integration of structured formats - which repositories are already connected and continuously provide new experiments for the repository? How many experiments are added every year/month/...?"
  "The section 4.1.2 decribes a lot of difficulties in matching experimental data with different parameters and levels of granularity. However, it is not clear what is managed by the system at the moment. The sentence in 4.1.3 about "handling volumes of data largely beyond those manageable manually" hints that the problem is solved to the certain extent, and the screenshots in Fig 3-6 suggest this too."
  "4.1.4 and managing changes in the models describe a very interesting problem of reusing the existing experimental results to predict the new ones or to get new results by re-running only some parts of the experiments. It is not clear, though, which use cases are supported at the moment and what limitations the system still has."
  "on p.17, section 5.4 it is not clear who performs manual interventions (if at all). Or is it one more "possibility" for the system rather than implemented feature?"
- **Response**: We clarified what has been developed (or not developed) in Section 4.2. In general, Section 4.1 describes the use cases which constitute the goals of the current work, while in Section 4.2 we

describe the prototype listing the features already developed. Several use cases require to face challenges which represent the ongoing research, and constitute the requirements described in Section 5.
Regarding the upload of new experiments by domain experts, the feature is currently supported and a new figure with the related screenshot from the prototype has been added. However, we clarified better that there is not a direct integration with existing repositories currently, even if there are functions to batch-import data from third-party formats. Re-running only a subset of the experimental dataset is currently supported if done manually (through a "search and execute" process), however the automatic identification of the partial set is a future research goal and part of the envisioned system.

- **Reviewer**: "Section 6 should correspond to 4-5. Right now it describes "continuous data integration" which is only explained at high-level in 4 and 5. The same concern relates to the discussion of using the domain-ontologies vs OLAP - it would be nice to see in 4 and 5 how and what is used, then such choice becomes clearer."
- **Response**: Section 6 has been reviewed and expanded. In its new version, it describes the designed architecture and not the developed features, which have been described together with the prototype in Subsection 4.2. Following the suggestion, we added a description of data management in the current prototype in Subsection 4.2, before talking about the general data model in the architecture section.

- **Reviewer**: "In the description of the scientific model (p.19) and also earlier the authors mention 10^5-10^6" experimental parameters. It is not clear how the user can search with so many parameters (Fig 3) then. Could you please explain this better?"
- **Response**: Following the suggestion, that part has been clarified.
- 10^5-10^6 (the number is omitted in the new version) are the model parameters, i.e. the parameters that can be tuned by domain experts to change a model behavior. These parameters are currently not seen by the framework which simply "runs" a model as a black-box and gets its results. That paragraph discusses the idea of integrating the internal information of a model (and, therefore, also its parameters) in the framework data model to enable more complex analysis (for example, relating some model internal changes to some performance changes).

Review #3

- **Reviewer:** "##Overview

  The main value I find in this article is that it identifies and describes well requirements for experiment-based model development, and in particular showing the issues that must be addressed when automating and scaling up such research across multiple open data sources. As I think this would apply across domains, I would have liked some citation to similar work in automating modelling work for other fields, for instance in systems biology.

  I think this paper should be accepted following a minor revision. Some more concern must be placed on the language.

  A detailed annotated PDF is attached in the web version of this review at"
- **Response:** Following the suggestion, we described other examples of domains which share the same challenges, including systems biology, at the end of Section 3. We also generalized the Introduction section to be less domain-specific.
  We included all the corrections attached in the PDF. We thank the Reviewer for all the detailed comments and useful suggestions.

- **Reviewer**: "## Architecture section

The wording in the section of "Proposed Architecture" is floating between describing a potential general architecture ("It could be translated in the future") and features of the existing developed prototype ("the database has been designed..to privilege performance").

While I can read between the lines that the architecture was partially derived from the development of the prototype (which is good), this section attempts to give the opposite picture. This means a tension is artificially introduced that confuses the reader as to what parts of the architecture has been realized or not.

I suggest to be more concrete in the architecture section and focus on what has been implemented. The other design ideas are well reasoned and should be kept, but I would move them to a new subsection on future architectural work. This will show more clearly the distinction between features you can prove with the prototype and potential benefits which implementation (e.g exploratory OLAPs) may have hidden pitfalls yet to be discovered.

I have some questions on the choice of Service-Oriented Architecture. I understand the authors wanted to support multiple modelling systems and data formats, and so argue that individual workflow functions should be services to facilitate interoperability. While I certainly recognize this reasoning (as a developer of the Web Service-based workflow system Apache Taverna), I would also disagree with the argument that simply using SOA means data interoperability is easy."

- **Response**: We heavily improved the paper with respect to this point. In particular, what the reviewer pointed out: "while I can read between the lines that the architecture was partially derived from the development of the prototype (which is good), this section attempts to give the opposite picture" was certainly true, because we followed an agile development cycle and both the architecture and the requirements have been refined during the development of the prototype, but the previous version of the paper gave the impression that the general architecture was only a the starting point.
  Therefore, we now clarified the fact that both the architecture and the requirements are also the result of the prototype development.
  Moreover, following the suggestion, we divided already developed features from design proposals. We included all and only the features already developed in the prototype subsection (4.2), while the architecture section (6) put them in the general perspective of the overall design also with not implemented components.
  Doing this, the architecture section is now both the result of the prototype but also its future development, as described at the beginning of the section:
  "As illustrated previously in Subsection 4.2, some of the components and features have been already developed in the prototype. The following architecture includes, besides them, other components and features that must be considered as not developed yet. The development of the existing features in the prototype has proved to be crucial to refine the following architecture."

  We agree with the reviewer that simply using SOA does not automatically translate in an easy interoperability. We extended that part and clarified that our goal is to obtain interoperability (but also scalability, integration and independent development) mainly through the combination of two factors:
    - The usage of independent services
    - A lightweight communication layer among the services, which is constituted by the operational database. This database is designed to store very general and service-independent results (basically, it can store generic name-results pairs without semantic constraints on the results).

  Doing so, each service can autonomously produce results associated to some entities (an experiment, a set of experiments, etc.) and each other service can retrieve the results by name. Data curation services can ensure data quality and management on the stored results. In this regard, our proposal is more similar to a microservice architecture.
  Anyway, we agree that features such as interoperability are not guaranteed only by some architectural choices and we clarified our intentions.
  We also added a reference to Apache Taverna in the related work discussing WMS solutions.

- **Reviewer**: “## (Lack of) availability

  The article focuses for a large part around development of a prototype. Yet, this prototype seem *not* to be available except for a couple of screenshots.

  From

  > All relevant data that were used or produced for conducting the work presented in a paper must be made FAIR and compliant with the PLOS data availability guidelines prior to submission.

  In addition to a URL, I would highly recommend the authors to provide Open Source code of the developed prototype.

  An associated Zenodo DOI can then be used as a Code Citation from the paper.”
- **Response**: The source code of the prototype has been made available under the MIT License, together with some sample data, on https://github.com/sciexpem/sciexpem
  However, we did not associate a DOI to our code because Zenodo requires to do a release and archive the code to associate a DOI.
  In our case, the code cannot be considered ready for a release, since it is a prototype under active development and contains also partially developed and demo features.
  In this paper, the main goals of the prototype are requirement elicitation, refinement of the architecture, and improved presentation of some of the designed features, but is not meant to constitute a contribution per se.
  The paper presented at SAVE-SD sketched requirements and architecture starting only from the limitations of previously developed software. This paper heavily improved them, and the prototype has been an important tool to this goal. We better clarified this throughout the paper.