Dear Drs.
Alberto Laender,
Mirella Moro,
Vanessa Braganholo,
Journal of Information and Data Management Editors,

Please find attached the revised version our paper entitled "*Extracting and Semantically Integrating Implicit Schemas from Multiple Spreadsheets of Biology based on the Recognition of their Nature*", submitted for consideration for the special section on invited SBBD 2012 short papers of *JOURNAL OF INFORMATION AND DATA MANAGEMENT*.

We would like to thank the reviewers for their helpful comments and suggestions. We covered the issues raised by the reviewers in this new version. We clarified that, although our system has been conceived to be generalizable to other domains, it was implemented and validated only in the biology domain. Thus, the title of the paper was modified to restrict the scope and the text of the paper makes explicit that, in this moment, this research is focusing in the biology domain. A new explanation of the "spreadsheet nature" and its role in the schema recognition was introduced to make clearer our strategy. Due to some questions concerning integration of instances, we clarified in several parts of the text that our focus and contribution address implicit schemas recognition and integration. Our approach to integrate instances applies existing algorithms, but takes advantage of the recognized schemas. We also thoroughly checked the paper for English language mistakes and rephrased parts of the text, further improving readability.

Please find below our response to the reviewers. Each point raised is individually discussed here to facilitate tracking the changes in the paper. We look forward to hearing back from you.

Best regards,
Ivelize Rocha Bernardo
Matheus Silva Mota
André Santanchè

Review Points

Reviewer #1

Issue 1.
In this new version it became more clear that your system is heavily domain dependent. In all steps you exploit features that are very specific from the biology domain. This is not a weakness, since this is a very important domain to address. However, the authors should state up-front this bias instead of trying to sell their contribution as general. If it is not the case and I am wrong, the authors should present experiments in other domain.

Response.
Even though our approach was conceived to be generalizable, our implementation and tests are, in fact, specific in the biology domain. We modified several points in the text, including the title, to make more explicit that this recognition is to biology domain, as requested.

Issue 2.
A critical example of this too ambitious claim is trying to say that your system is able to "recognise the nature" of the spreadsheet, while it is now clear that it in fact recognises if the spreadsheet is likely to contain data from the biology domain or not, relying on a crafted set of heuristics based on patterns. It should be interesting if the authors could better characterise their contributions to the benefit of the interested readers of JIDM.

Response.
Besides the ability of recognizing if the spreadsheet contains data from the biology, as observed, the spreadsheet nature involves recognizing a construction pattern associated to its domain. This pattern reflects a conceptual model hidden in the implicit schema of the spreadsheet, and is exploited to drive the recognition of the fields and their relations. We added a paragraph in the introduction, plus extra explanations throughout the text, to make the notion of "spreadsheet nature" clearer, as well as how we recognize and exploit it.

Issue 3.
I believe you misunderstood my requirement regarding experimentation. To properly evaluating your work, it is important to asses the quality of the results achieved in all tasks your system carries out. You have only shown numbers on the first task, which is "Schema Recognition". From my point of view, the authors are required to present an evaluation of the other tasks, there is "Instance Mapping" and "Semantic Enrichment" and "Data Integration". These tasks are considerably more complex then first one and it is mandatory to provide evidences of the effectiveness of your system in carrying out all of them.

Response.
The focus of our differential (the implicit schemas and not the instances) was not clear enough and consequently it created an expectation beyond our scope. The

integration of instances in this work appears as consequence of the integration of schemas. We applied existing matching algorithms to integrate instances and, therefore, it does not bring a novelty by itself. For this reason, we did not present assessments in these steps. Nevertheless, our intention in presenting a practical example involving integration of instances was to show the potential of integrating schemas in this process.

To clarify our differential and contribution, we included several extra explanations in the: introduction, comparison with related work and in the explanation of our proposal (see a detailed explanation in Section: Semantic Enrichment). In a nutshell, we make clear that our focus and contribution addresses the schema recognition and semantic mapping, explaining the role of instance recognition and mapping in the scenario.

Reviewer #2

Issue 1.
In the new version, there have been remarkable improvements on the structure and contents of the article. However, it still needs a careful review on its text. I have made several comments along the text (please see the attached file at the JIDM site), but given the number of mistakes, the authors should double check for the correct use of the language.

Response.
We thoroughly checked the paper for English language mistakes and rephrased parts of the text. We reorganized sections and renamed some of them.

We applied all the corrections suggested by the reviewer in the attached file.

Thank you very much for the detailed work done.