Author's response to reviews

Title: Ancillary Study Management Systems: A Review of Needs

Authors:

Elizabeth K Nelson (eknelson@labkey.com)
Britt Piehler (brittp@labkey.com)
Adam Rauch (adam@labkey.com)
Sarah Ramsay (sramsay@scharp.org)
Drienna Holman (drienna@scharp.org)
Smita Asare (sasare@immunetolerance.org)
Adam Asare (aasare@immunetolerance.org)
Mark Igra (marki@labkey.com)

Version: 2 Date: 22 October 2012

Author's response to reviews: see over
Editors, BMC Medical Informatics and Decision Making
BioMed Central Ltd
Floor 6
236 Gray's Inn Road
London WC1X 8HL
United Kingdom

Dear Editors and Reviewers,

Thank you for your many thoughtful suggestions for improving our paper, *Ancillary Study Management Systems: Review of Needs and Options*. Your detailed suggestions have helped us significantly enhance the manuscript. Please consider our revised paper for publication as a debate article in *BMC Medical Informatics and Decision Making*.

To address the many helpful comments we received, we have:

- Addressed issues surrounding system interoperability, integration, and data exchange, plus the use of standards to tackle these challenges
- Explained the derivation of the term “ancillary study data management system” (ASMS)
- Incorporated additional references to provide broader, deeper context for our discussion of ASMSs
- Added figures to illustrate our conceptual model for the design and use of an ASMS:
  - Figure 1 - a proposed set of ASMS data flows for a common workflow
  - Figure 2 - a visualization of how an ASMS may facilitate “virtuous cycles” of information review and reuse
- Updated the title and text to clarify that the paper is intended as an analysis of scenarios and requirements, not as a report on one specific software system
- Explained our methodology more thoroughly
- Fixed a variety of smaller format, content and wording issues
- Added a list of the abbreviations
Detailed responses to reviewer suggestions are provided on subsequent pages.

All authors have had the opportunity to review the revisions and contribute edits. All authors approved the text.

Thank you again for your time and thoughtfulness in reviewing our paper.

Sincerely,

Elizabeth Nelson
Review #1

Reviewer: Philip Payne
Reviewer’s report:

1. This report describes a set of information needs related to the design and utilization of what the authors describe as “Ancillary Study Management Systems” (ASMS)– which is defined in this particular report as being as informatics platforms that facilitates the secondary use of data for investigational purposes, where such data are to be derived from the primary or synthesized data generated during the course of other studies or programs of research. At a high level, the rationale for why such systems are needed is both interesting and novel, and is well stated by the authors. Furthermore, the authors provide a well-articulated set of information needs related to their focus area, thus supporting the generalizability and extensibility of their work. This reviewer believes that this report will ultimately be of great interest to the clinical research informatics community, but does have a number of deficiencies that should be addressed prior to publication, as follows:

REPLY: We thank Dr. Payne for his enthusiasm for the level of interest of this work and for his insightful suggestions for fixing deficiencies.

Major Compulsory Revisions:

2. While the authors describe a succinct and at a surface level, rational set of information needs related to the design and use of ASMS, the report does not provide any details concerning how such information needs were systemically defined, verified, or validated. Such methodological rigor would greatly enhance the impact of this submission.

REPLY: We added a paragraph to the Methods section to provide more detail on our methodology:

The process we used to define and verify ancillary study management scenarios and requirements was deliberatively iterative. We conducted repeated interviews with collaborators working with representative organizations, particularly CHAVI, ITN, HVTN, MHRP, SCHARP, HPTN and MTN. We provided written copies of the scenarios and requirements distilled from these interviews to interviewees at these organizations. After individuals reviewed these written materials, we conducted follow-up interviews and updated the written scenarios and requirements to reflect detailed feedback. The ITN, HVTN, MHRP and SCHARP collaborators participating as co-authors provided written sign-off on the final set of scenarios and requirements. We are concretely validating these scenarios and requirements as we develop the ancillary study management
features of LabKey Server and verify these features with researchers using the system to conduct ancillary studies.

*We acknowledge that this study did not include a formal survey (in the vein of Anderson et al. 2007), so our methods are not as quantitative as would be ideal. Also, we acknowledge that our interviews were not executed, recorded and analyzed as formally as Embi and Payne’s 2009 CRI survey.*

3. In the review of pertinent background materials that contributed to the authors understanding of the field of clinical research informatics, especially as it pertains to the design and functionality of clinical research management systems or equivalent informatics platforms, a significant number of notable reports from authors such as Chung, Embi, Richesson, Payne, Weng, and others is omitted. A more comprehensive set of references should be included in this review.

*REPLY: Indeed, many more articles informed our understanding of clinical management systems than the ones cited in the original article. You are correct that the original article leaned too far towards brevity in citations. We hope that we have expanded our citations in a manner that better balances depth with brevity.*

A. *Modifications to the Background section are underlined:*

….Searching PubMed for the term “ancillary study” and its plural produces 889 results ([10], October 2012). A recent review of the emerging field of clinical research informatics identifies secondary data use as a key pain point([11, 12]. From a broader perspective, reuse of scientific data has become a priority and a concern across fields of all science, not just biomedical research studies([13, 14].

For primary studies and trials, clinical trial management systems (CTMSs) are widely used to enhance efficiency, reduce costs, comply with regulations, and speed up data analysis. Today, researchers have a wide range of CTMS options at their disposal, from proprietary solutions such as Oracle Corporation’s Oracle Clinical([15] and Phase Forward Clintrial([16] to open source solutions such as TrialDB([17] and OpenClinica([18, 19]. CTMSs are actively studied and developed by the academic community([20–24].

In contrast, with the exception of our own conference abstract([25], the research literature contains little discussion of systems (or system extensions) that specifically address the cradle-to-grave needs of ancillary studies. The open source i2b2 system([26–28] is particularly noteworthy for its support for de-identified cohort discovery across federated patient information repositories. However, descriptions of i2b2 focus primarily on the repurposing of information and material by-products of health care delivery([26–28], not on the full set of scenarios surrounding reuse of the products of trials and studies.
New references in the paragraphs above include:


B. Articles by Richesson and others regarding standards have been cited in the Discussion section:

Standards relevant to some of these data transfer pain points are already available (e.g., standards for CRF results[69]) while others are undergoing ongoing active development[70–74].

Citations in the sentence above now include:


C. Payne et al.’s 2010 paper on TRITON should have been further cited in the Discussion section’s mention of the push towards open source systems. It is now cited there:

Given the calls for open source CTMSs in the recent past[19, 24], we expect open source ASMSs to be desirable.

D. The discussion now calls out a recent analysis of motivators for adoption of biomedical research systems:

Our experience squares with a recently published comparison of the success factors for REDCap and i2b2 and the hindrances of caBIG[82].

The relevant citation:


E. Figure 2 now conceptually and specifically references Kahn and Weng 2012, as underlined below. (Full text of Figure 2’s caption is included below, in response to item #4.)
...(i) Full study cycle. The nine steps in the ancillary study workflow described in this paper form a virtuous cycle that spans the full life of a study, from the first glimmer of an idea through publication. For simplicity, Figure 2 breaks these steps into three phases (study initiation, study execution, and results sharing). The steps in this cycle are roughly equivalent to those in the “inner,” study-based loop in Kahn and Weng’s conceptual model for clinical research informatics[83]. In such cycles, published hypotheses and shared data from completed studies are used to generate future discovery cycles by providing inspirations and ingredients for follow-up studies.

The relevant citation:


4. While the authors present their information needs as distinct functional requirements, there appears to be an underlying conceptual model for the design and use of ASMS that is implicitly being used to inform such analyses and discourse. This model should be formalized, and reported, as an organizing framework for the information needs. This would likely involve some type of visual representation of that conceptual model.

REPLY: Excellent point, one made by Reviewer #2 as well. We added two figures, one showing our conceptual models for ASMS data flows and a second showing two kinds of “virtuous cycle” feedback loops that can be facilitated by an ASMS.

Figure 1 - ASMS data flows

This figure shows a conceptual model for data flows for ancillary studies whose primary focus is analysis of stored specimens. In this scenario, data flows into the ASMS from the primary study’s CTMS (which contains information on participants, visits, consent and other pre-existing data) and specimen repository LIMS (Laboratory Information Management Systems, which contain information on stored specimens available for further investigation). Before an ancillary study is initiated, the ASMS is used for hypothesis generation and feasibility investigations based on specimen availability. Once a particular ancillary study has been identified, a container for its data is established within the ASMS. After the ancillary study has been approved, any additional participant consents required for the study are collected by clinical sites and noted in the ASMS. Requests for needed specimens (including material transfer agreements) are sent to the appropriate specimen repositories, which in turn send stored specimens to appropriate labs. The labs perform assays on the specimens and import the results to the ancillary study container in the ASMS. Once the ancillary study is complete, results may be repatriated to the primary study. Results may also be shared in publications or other venues.
This model presumes that all data for the ancillary study is managed within the ASMS, not the CTMS or an external system. It also presumes that external investigators can be given access to the study within the ASMS. Under different assumptions, usage patterns and data flows would change, but an ASMS could still prove helpful.

For example, if gathering new clinical data from study participants is a significant piece of an ancillary study, using an organization’s existing CTMS for collecting and managing clinical data might make most sense. An ASMS could still be desirable for other aspects of the study. For our collaborators, CTMSs have not proven amenable to the kinds of queries necessary for hypothesis generation and participant identification. Also, they are not ordinarily well-integrated with relevant LIMS, so they do not facilitate identification of specimen availability. An ASMS could be used for these steps and others that are not typically supported by CTMSs or LIMS, such as specimen requests and assay data management.

![Figure 2 - Virtuous cycle feedback loops](image)

Using existing results and materials to refine hypothesis and develop new insights can produce “virtuous cycles” where the research efforts of today feed tomorrow’s discoveries. Figure 2 shows two kinds of such cycles that are implied by the ancillary study workflows discussed here. An ASMS can facilitate both types of cycles by smoothing the flow of information, enabling collaboration, simplifying workflows and allowing researchers to make the most of existing materials and information.

(i) Full study cycle. The nine steps in the ancillary study workflow described in this paper form a virtuous cycle that spans the full life of a study, from the first glimmer of an idea through publication. For simplicity, Figure 2 breaks these steps into three phases (study initiation, study execution, and results sharing).
The steps in this cycle are roughly equivalent to those in the “inner,” study-based loop in Kahn and Weng’s conceptual model for clinical research informatics[82]. In such cycles, published hypotheses and shared data from completed studies are used to generate future discovery cycles by providing inspirations and ingredients for follow-up studies.

(2) Incremental review cycles. An ASMS can also facilitate smaller-scale virtuous cycles during all phases of an ancillary study. First and foremost, during the study initiation phase, the information and tools made available by an ASMS allow incremental refinement of hypotheses and study plans according to existing data, specimen availability, and consent limitations. During later phases of a study, an ASMS can make it easier to share and review new information as it is collected, allowing feedback of new insights into study investigations, operations, analyses, and conclusions. Of course, in-progress studies governed by clinical trial regulations will provide less scope for immediate use of this type of feedback than the kinds of pre-clinical, exploratory studies common among our investigators.

5. Minor Compulsory Revisions:
Figure captions could be extended to provide comprehensive descriptions of the purpose of such visuals and describing key, salient features.
**REPLY:** For the new figures added as part of this revision, we provide more comprehensive captions. These captions are shown above. We cut the original document’s only figure (Figure 1, a LabKey Server screenshot) due to the concerns of Reviewer #3.

6. It is not clear if the term “Ancillary Study Management Systems” is widely used by the CRI community. The derivation, definition, and use of this term should be explained more clearly in the manuscript.

**REPLY:** The Results section now includes a section that answers these good questions:

**Naming considerations**

No genuinely standard terminology exists to describe systems for primary study management, let alone ancillary study management. For example, the term Clinical Data Management System (CDMS) is sometimes used as a synonym for CTMS (e.g., Ohmann et al. [72]). Given this variability, we made a best approximation when we chose the term ASMS. In our own experience, the key use case for such a system is management of data for follow-up studies that require analysis of stored specimens. However, we define the term ASMS more broadly, encompassing both secondary data analysis studies and studies that require additional measurements (clinical, specimen-based or both). We expect an ASMS to be useful in these situations as well.

We derived the name Ancillary Study Management System based on common terminology and use cases. During our review of research network protocols, the term **ancillary studies** came up most frequently and specifically as the term for follow-up studies that encompassed more than just secondary data analysis. The appearance of the term **ancillary studies** in NIH grant announcement titles[2–9] and the frequency of this term in PubMed searches[10] (889 hits, as of October 2012) confirmed that it is in common use. We define an ASMS as a system for managing data for follow-up studies. Such studies require collection of additional measurements, usually from stored specimens, or secondary analysis of data.
Review #2

Reviewer: Meredith Nahm
Reviewer’s report:

1. Major Comp. Revisions:
- The scenarios were helpful, but from a formal requirements gathering and specification process and for publication, I would expect to see more detailed artifacts such as a use case diagram, functional specifications, static (information model) and dynamic models (activity and state diagram) and a data flow diagram. If the journal supports electronic supplemental information, I would want to see this type of information along with the article.

REPLY: A good point, one made by Reviewer #1 as well. As shown below in Figures 1 and 2, we added diagrams to visualize our conceptual models for data flows and “virtuous cycle” feedback loops. We did not add additional diagrams for a variety of reasons, as described below. Regardless of these reasons, we recognize that the paper could have been improved by additional visualizations.

During development of the original paper, we prepared a basic dynamic model diagram. However, as part of our iterative interview process, we encountered variability between organizations regarding the details of this diagram. As a result, this diagram was repeatedly simplified to the point where it no longer provided clear value. We felt that publishing an activity diagram with sufficient detail to be interesting could be misleading or confusing. The consistent patterns and variations we encountered were best embodied in the workflow discussion provided in the results section. Nevertheless, we added Figures 1 and 2 to do a better job of visualizing these high-level patterns. Given our experience with the dynamic model diagram, we did not create a formal use case diagram for the original version of this paper.

We did not add functional specifications because we aimed to focus this paper on a review of scenarios and requirements, not on the development and features of a particular software system. Reviewer #3’s comments led us to further de-emphasize software development in order to avoid any confusion about the purpose of this paper.

Figure 1 - ASMS data flows
This figure shows a conceptual model for data flows for ancillary studies whose primary focus is analysis of stored specimens. In this scenario, data flows into the ASMS from the primary study’s CTMS (which contains information on participants, visits, consent and other pre-existing data) and specimen repository LIMS (Laboratory Information Management Systems, which contain information on stored specimens available for further investigation). Before an ancillary study is initiated, the ASMS is used for hypothesis generation and feasibility
investigations based on specimen availability. Once a particular ancillary study has been identified, a container for its data is established within the ASMS. After the ancillary study has been approved, any additional participant consents required for the study are collected by clinical sites and noted in the ASMS. Requests for needed specimens (including material transfer agreements) are sent to the appropriate specimen repositories, which in turn send stored specimens to appropriate labs. The labs perform assays on the specimens and import the results to the ancillary study container in the ASMS. Once the ancillary study is complete, results may be repatriated to the primary study. Results may also be shared in publications or other venues.

This model presumes that all data for the ancillary study is managed within the ASMS, not the CTMS or an external system. It also presumes that external investigators can be given access to the study within the ASMS. Under different assumptions, usage patterns and data flows would change, but an ASMS could still prove helpful.

For example, if gathering new clinical data from study participants is a significant piece of an ancillary study, using an organization’s existing CTMS for collecting and managing clinical data might make most sense. An ASMS could still be desirable for other aspects of the study. For our collaborators, CTMSs have not proven amenable to the kinds of queries necessary for hypothesis generation and participant identification. Also, they are not ordinarily well-integrated with relevant LIMS, so they do not facilitate identification of specimen availability. An ASMS could be used for these steps and others that are not typically supported by CTMSs or LIMS, such as specimen requests and assay data management.

Figure 2 - Virtuous cycle feedback loops
Using existing results and materials to refine hypothesis and develop new insights can produce “virtuous cycles” where the research efforts of today feed...
tomorrow’s discoveries. Figure 2 shows two kinds of such cycles that are implied by the ancillary study workflows discussed here. An ASMS can facilitate both types of cycles by smoothing the flow of information, enabling collaboration, simplifying workflows and allowing researchers to make the most of existing materials and information.

(i) **Full study cycle.** The nine steps in the ancillary study workflow described in this paper form a virtuous cycle that spans the full life of a study, from the first glimmer of an idea through publication. For simplicity, Figure 2 breaks these steps into three phases (study initiation, study execution, and results sharing). The steps in this cycle are roughly equivalent to those in the “inner,” study-based loop in Kahn and Weng’s conceptual model for clinical research informatics[82]. In such cycles, published hypotheses and shared data from completed studies are used to generate future discovery cycles by providing inspirations and ingredients for follow-up studies.

(2) **Incremental review cycles.** An ASMS can also facilitate smaller-scale virtuous cycles during all phases of an ancillary study. First and foremost, during the study initiation phase, the information and tools made available by an ASMS allow incremental refinement of hypotheses and study plans according to existing data, specimen availability, and consent limitations. During later phases of a study, an ASMS can make it easier to share and review new information as it is collected, allowing feedback of new insights into study investigations, operations, analyses, and conclusions. Of course, in-progress studies governed by clinical trial regulations will provide less scope for immediate use of this type of feedback than the kinds of pre-clinical, exploratory studies common among our investigators.
2. For a complimentary system such as that favored in the manuscript, I would expect and want to see a lot more information about interoperability. For example, interface with the CTMS, electronic IRB, and LIMS systems. There are standards available for some of these, and standards are lacking for others. A thorough analysis of where the software can automate workflow through supporting available standards should be a key part of a requirements analysis. An analysis of where standards are lacking and what the research team proposes to do about it, i.e., propose and/or work toward the needed standards, how to fill the gap until the needed standards are available.

REPLY: Another excellent point and a clear weakness of the original manuscript. We largely reworked the Discussion section to improve our discussion of the challenge of interoperability and the importance of standards.

Two notes:
- We chose to address standards primarily in the Discussion section instead of the requirements section in order to avoid confusion between the potential benefits of the using of an ASMS and the use of standards.
- Reviewing available standards in the interface-by-interface manner suggested would take a greater degree of expertise in standards than we have at this time. Rather than overreach, we hope that the perspective on the use of standards that we provide based on our experience as system
developers may still be valuable. If it is not, we would be happy to remove it. We made sure to include references to more authoritative articles on standards relevant to ancillary study data management.

Discussion

Opportunities and challenges for ASMS use

ASMSs designed around the scenarios described here can help research organizations overcome the operational and data management challenges of ancillary studies. By providing tools for difficult, repetitive or complex tasks, ASMSs can lower costs, increase efficiencies, produce more accurate results, expand the use of valuable primary study data, and reduce both expertise and personnel requirements. For organizations with extensive existing software systems, introducing an additional data management system may pose substantial challenges in integration, interoperability, data redundancy and data exchange. This may eventually motivate incorporation of ASMS features into CTMS systems; however, having a separate, complementary ASMS may still make sense for certain organizations.

Our collaborators exemplify users who have found value in developing ASMSs that are complementary to existing systems. Key motivators for system development have been hypothesis generation and specimen allocation for pre-clinical, not clinical, studies. Their questions have been exploratory and outside of the FDA pipeline, so they have not needed or wanted the overhead of working within a fully validated CTMS. Furthermore, their in-house software infrastructure did not provide natural points for integrating ASMS features. Some had arms-length relationships with the CROs (contract research organizations) running relevant CTMSs, while others did not employ comprehensive, integrated CTMSs. For this reason, direct interoperability of the ASMS system with a CTMS was not desired or needed.

Even when integration of ASMS features into a CTMS may be ideal, it may not be practical, so a complementary ASMS may be desirable. For organizations that have a CTMS, updates to CTMS systems can be nontrivial, particularly when system validation is required for clinical studies. Furthermore, supporting the specimen allocation scenarios of ancillary studies requires some degree of integration with one or several LIMS, which may not be feasible for a CTMS.

If an ASMS is used as a complement to existing systems, interoperability, data exchange and data integration may be somewhat eased by the further development and adoption of standards, including standardized ontologies. The arrows in Figure 1 suggest many of the potential pain points for data exchange in the workflows discussed here, including the transmission of: (1) primary study results from CTMS to ASMS, (2) specimen information from specimen repositories’ LIMS to ASMS, (3) consent information from clinical sites to ASMS or CTMS (4) specimen requests (including MTAs) from ASMS to specimen repositories, (5) specimen metadata from specimen repositories to assay labs, (6) assay results to ASMS, (7) ancillary study results from ASMS to CTMS, and (8) ancillary study results from ASMS to
publications or other venues for sharing results. Additional pain points that are not shown include the transfer of (1) case report form (CRF) results to CTMS and later to ASMS; (2) primary study consent and IRB information to CTMS and later to ASMS; and (3) ancillary study consent and IRB approval to ASMS; and (4) assay results from laboratory machines to lab record systems. Data transfer becomes further complicated when upstream errors occur and changes need to propagate to downstream systems, or when quality control mistakes are identified downstream and need to propagate upstream. Even when data exchange succeeds, semantic variations in delivered data can stymie comprehension and meaningful use.

Standards relevant to some of these data transfer and comprehension pain points are already available (e.g., standards for CRF results[69]), while others are undergoing continued, active development[70–74]. Widely used standards have the potential to facilitate the reuse of study data, the automation of ancillary study workflows, and even the reuse of software systems themselves. However, in our experience developing software systems, systematic use of standards is currently more the exception than the rule. For this reason, the adoption of standards cannot be assumed as a uniform solution for many of the data transfer challenges mentioned here. This may change as NIH grant requirements drive a greater degree of adherence to standards. For example, the recent funding opportunity announcements for the competitive renewal of the HVTN[75], HPTN[76], and MTN[77] require CDISC[78] compliance for the collection, storage and transfer of study data.

While standards do yet provide a complete solution, some degree of usage of both official standards (such as CDISC) and pervasive, industry-developed file formats can still reduce integration hurdles and foster system adoption. In our experience disseminating the LabKey Server open source platform, we have found that organizations that use data formats already incorporated in the platform (e.g., Frontier Science’s LDMS[79] format for describing specimens and Clinical DataFax System Inc.’s DataFax[80] format for defining datasets and visit maps) consider this a plus for adopting LabKey Server. LabKey Server itself emphasizes well-documented archive formats[81] that convey study data via tab-separated value text files and metadata, properties, and settings via simple XML files. A variety of organizations (e.g., SCHARP) use these formats to integrate external study data into their LabKey Server systems. However, we have seen few industry-defined data formats (or official, community-defined standards) that are sufficiently widely used to be significant drivers of adoption and use. Instead, providing flexibility in accommodating an organization’s existing or newly invented data types (e.g., LabKey Server’s graphical wizard for describing new assay data types[32]) has proven far more important to adoption. Our experience squares with a recently published comparison of the success factors for REDCap and i2b2 and the hindrances of caBIG[82].

The requirements analysis section also had a pre-existing section that briefly touches on standards for consent management. We emphasized this particular
standards issue in the requirements section because of the frequency we heard it mentioned during our interviews:

**Key Scenarios for Ancillary Study Management**

4. *Obtaining or verifying consent.* …

Electronic management of consent is highly desired for all types of studies, not just ancillary studies[64], as are ontology-based definition of consent[65, 66] and collection of metadata about consent coverage. These enable standardization of consent across network trials; reduce ambiguity; allow more efficient and consistent determination of consent coverage and status; increase the likelihood that specimens and data can be re-used in future studies; and facilitate compliance with consent agreements that require the destruction of participant specimens at a particular time or upon participant withdrawal from any phase of a trial.
Review #3

Reviewer: Wenle Zhao
Reviewer’s report:

Major Compulsory Revisions:
1. The manuscript discussed the information management system requirements for ancillary study management. This topic is of high importance in medical researches.

REPLY: No changes are suggested in this item.

2. Based on the contents of the manuscript, I had the impression that a computerized database system has been developed and implemented for the management of actual ancillary studies. However, throughout the article, no information on the design of architecture structure of such system was shared with the readers. The scenarios and requirements discussed in this manuscript are important for the planning of the ASMS. But they are far from sufficient.

REPLY: We understand and appreciate your desire for additional information on LabKey Server. However, this paper aims to explore scenarios and requirements of general interest, not the specific features and architecture of the LabKey Server system. With the exception of the ancillary study features currently under development, LabKey Server’s features and architecture are covered fully elsewhere. Relevant references (Nelson et al. 2011, Piehler et al. 2011 and Rauch et al. 2006) are cited in the paper. We aim to cover LabKey Server’s ancillary study features in a future paper, when they are ready.

To clarify the focus of the paper, we have made multiple changes.

A. The title has been changed from:
   Ancillary Study Management Systems: A Review of Needs and Options
   To:
   Ancillary Study Management Systems: A Review of Needs

B. The “Current development efforts” section of the Discussion includes several clarifying comments:

   Current development efforts
   In collaboration with CHAVI and the ITN, we are currently developing a general-purpose, open-source ASMS based on the LabKey Server system. We mention this system only briefly to encourage participation an open source project that aims to gradually address ancillary study scenarios. This paper does not explore the ancillary study features of this system because these features are currently under development.
…The basic features and architecture of LabKey Server are explored elsewhere[32, 34, 35].

C. The former Figure 1, which showed a screen shot of LabKey Server, has been removed to further de-emphasize the software.

3. It is unclear that the ASMS mentioned in this article is a computerized database information system, like those well-known Clinical Trial Management System (CTMS) or Electronic Data Capture (EDC) system with Graphical User Interfaces (GUI), a computerized program, or a standardized operation procedure (SOP)?

*REPLY*: The Discussion section now provides this brief note on the architecture of LabKey Server, which the citations elaborate (e.g., Nelson et al. 2011):

LabKey Server is a web application implemented in Java that runs on the Apache Tomcat web server and stores its data in a relational database engine. … The basic features and architecture of LabKey Server are explored elsewhere[32, 34, 35].

4. If the ASMS has been actually developed, I feel it will be necessary to provide key design features to the readers, and also answer the questions the author challenged existing alternatives (described in pages from 9 to 12). For example, how to link to existing CTMS? How to work with on-going studies? How to avoid redundant data capture activities on regulatory document, clinical data, study progress data? How to manage user accounts, user access permission? How to protect data safety and integrity? How to control data transfer from the primary CTMS based on IC status for the ancillary study? Who is responsible for data queries raised from ancillary study investigators? How to handle data error when the primary study data is locked?

*REPLY*: As mentioned in our reply to item #2, this paper does not aim to cover the features and implementation of LabKey Server, nor any particular instance of LabKey Server or an ASMS.

You are quite right that concerns about interoperability and redundancy were not fully emphasized in the original paper. These are now more fully emphasized in the Discussion section:

**Discussion**

**Opportunities and challenges for ASMS use**

ASMSs designed around the scenarios described here can help research organizations overcome the operational and data management challenges of ancillary studies. By
providing tools for difficult, repetitive or complex tasks, ASMSs can lower costs, increase efficiencies, produce more accurate results, expand the use of valuable primary study data, and reduce both expertise and personnel requirements. For organizations with extensive existing software systems, introducing an additional data management system may pose substantial challenges in integration, interoperability, data redundancy and data exchange. This may eventually motivate incorporation of ASMS features into CTMS systems; however, having a separate, complementary ASMS may still make sense for certain organizations.

Our collaborators exemplify users who have found value in developing ASMSs that are complementary to existing systems. Key motivators for system development have been hypothesis generation and specimen allocation for pre-clinical, not clinical, studies. Their questions have been exploratory and outside of the FDA pipeline, so they have not needed or wanted the overhead of working within a fully validated CTMS. Furthermore, their in-house software infrastructure did not provide natural points for integrating ASMS features. Some had arms-length relationships with the CROs (contract research organizations) running relevant CTMSs, while others did not employ comprehensive, integrated CTMSs. For this reason, direct interoperability of the ASMS system with a CTMS was not desired or needed.

Even when integration of ASMS features into a CTMS may be ideal, it may not be practical, so a complementary ASMS may be desirable. For organizations that have a CTMS, updates to CTMS systems can be nontrivial, particularly when system validation is required for clinical studies. Furthermore, supporting the specimen allocation scenarios of ancillary studies requires some degree of integration with one or several LIMS, which may not be feasible for a CTMS.

If an ASMS is used as a complement to existing systems, interoperability, data exchange and data integration may be somewhat eased by the further development and adoption of standards, including standardized ontologies. The arrows in Figure 1 suggest many of the potential pain points for data exchange in the workflows discussed here, including the transmission of: (1) primary study results from CTMS to ASMS, (2) specimen information from specimen repositories’ LIMS to ASMS, (3) consent information from clinical sites to ASMS or CTMS (4) specimen requests (including MTAs) from ASMS to specimen repositories, (5) specimen metadata from specimen repositories to assay labs, (6) assay results to ASMS, (7) ancillary study results from ASMS to CTMS, and (8) ancillary study results from ASMS to publications or other venues for sharing results. Additional pain points that are not shown include the transfer of (1) case report form (CRF) results to CTMS and later to ASMS; (2) primary study consent and IRB information to CTMS and later to ASMS; and (3) ancillary study consent and IRB approval to ASMS; and (4) assay results from laboratory machines to lab record systems. Data transfer becomes further complicated when upstream errors occur and changes need to propagate to downstream systems, or when quality control mistakes are identified downstream and need to propagate upstream. Even when data exchange succeeds, semantic variations in delivered data can stymie comprehension and meaningful use.
Standards relevant to some of these data transfer and comprehension pain points are already available (e.g., standards for CRF results[69]), while others are undergoing continued, active development[70–74]. Widely used standards have the potential to facilitate the reuse of study data, the automation of ancillary study workflows, and even the reuse of software systems themselves. However, in our experience developing software systems, systematic use of standards is currently more the exception than the rule. For this reason, the adoption of standards cannot be assumed as a uniform solution for many of the data transfer challenges mentioned here. This may change as NIH grant requirements drive a greater degree of adherence to standards. For example, the recent funding opportunity announcements for the competitive renewal of the HVTN[75], HPTN[76], and MTN[77] require CDISC[78] compliance for the collection, storage and transfer of study data.

While standards do yet provide a complete solution, some degree of usage of both official standards (such as CDISC) and pervasive, industry-developed file formats can still reduce integration hurdles and foster system adoption. In our experience disseminating the LabKey Server open source platform, we have found that organizations that use data formats already incorporated in the platform (e.g., Frontier Science’s LDMS[79] format for describing specimens and Clinical DataFax System Inc.’s DataFax[80] format for defining datasets and visit maps) consider this a plus for adopting LabKey Server. LabKey Server itself emphasizes well-documented archive formats[81] that convey study data via tab-separated value text files and metadata, properties, and settings via simple XML files. A variety of organizations (e.g., SCHARP) use these formats to integrate external study data into their LabKey Server systems. However, we have seen few industry-defined data formats (or official, community-defined standards) that are sufficiently widely used to be significant drivers of adoption and use. Instead, providing flexibility in accommodating an organization’s existing or newly invented data types (e.g., LabKey Server’s graphical wizard for describing new assay data types[32]) has proven far more important to adoption. Our experience squares with a recently published comparison of the success factors for REDCap and i2b2 and the hindrances of caBIG[82].

5. A CTMS is a comprehensive information management system. It could cover many aspects of clinical trial managements. The challenges caused by unique requirements for ancillary studies may not necessary overweight the burden of creating a new information management system to cover the basic common requirements of clinical studies (both clinical trials and ancillary studies).

REPLY: The Discussion has been significantly revised to addresses the tradeoff of having a separate ASMS. This section is shown above, in response to item #4.
6. The author listed Excel, SAS as project and data management tools as alternatives to the proposed ASMS. In clinical research practice, Excel spreadsheets and SAS programs can’t be considered at the same level as a centralized, usually web-based, information system like CTMS. They are not on the same levels regarding of functionality scopes, reliability, and data safety.

REPLY: Indeed, you are quite correct. Using a combination of ad hoc software tools does not provide the same level of functionality, reliability, and data safety as using a centralized, web-based information system. However, the ad hoc strategy is discussed here because of the frequency of its use, as encountered during our interview process.

7. The readers may want to know the proposed ASMS is developed in what platform? Is it a Excel Spreadsheet manually managed by an investigator or it is an information management system with a 3-tire structure?

REPLY: The Discussion section now provides this brief note on the architecture of LabKey Server, which the citations elaborate (e.g., Nelson et al. 2011):

LabKey Server is a web application implemented in Java that runs on the Apache Tomcat web server and stores its data in a relational database engine. … The basic features and architecture of LabKey Server are explored elsewhere[32, 34, 35].

Minor Essential Revisions:
1. Missing page numbers.

REPLY: Page numbers have been added.

2. Confusing numbering of sections and paragraphs.

REPLY: We have revised the sections such that only two sections have items numbered. Numbers in these two sections are now differentiated; the first section uses Roman numerals, while the second uses Arabic numbers.

Details on the changes, listed by section title:
- All numbers cut: “Differentiating characteristics of ancillary studies”
- Numbered list now uses Roman numerals: “Existing alternatives for managing ancillary studies”
- Numbered list continues to use Arabic numbers: “Key Scenarios for Ancillary Study Management”
Reviewer: Greg Fegan
Reviewer's report:

1. General Comment
I believe that this paper makes a valid contribution in line with the journal’s guidelines and have made the following suggestions for amendments in a Discretionary Revision mode.

Does the debate present a novel argument, or a novel insight into existing work?
I think that the authors have presented a cogent outline of how they are planning, and what criteria they used, to add extra functionality and utility in an important direction to their existing system. They used a vignette approach ie a “freezer study” scenario that I believe many people will be familiar with.
Does the debate address an important problem of interest to a broad biomedical audience?
The use of data and its re-use and overall importance has started to gain more recognition in my opinion over the last few years and the authors of this piece intend to assist that process.

REPLY: No revisions requested in this section.

2. Is the piece well argued and referenced?
I believe that the authors can slightly strengthen their case by i) Giving more context to the primacy of data and data sharing which is what underlies their approach. For example, I believe that the Science February 11th 2011 special issue (see http://www.sciencemag.org/site/special/data/) is a great starting place.
The concluding line in the Editorial of this issue by the Science Editors where they state “We must all accept that science is data and that data are science, and thus provide for, and justify the need for the support of, much improved data curation.” is, in my opinion, likely to mark a quite profound change in data sharing in the near term.

REPLY: A good point. To address this, the first paragraph of the Background section now cites references from the field of biomedical informatics, as well as articles from the Science special issue you mentioned:

…. A recent review of the emerging field of clinical research informatics identifies secondary data use as a key pain point[11, 12]. From a broader perspective, reuse of scientific data has become a priority and a concern across fields of all science, not just biomedical research studies[13, 14].
3. ii) Where the authors cite themselves as in [19] and [23] I think it would be better to identify this as such by using a construct such as: “We have previously...”.

REPLY: Thank you for the thoughtful advice. Revisions:

…In contrast, with the exception of our own conference abstract[25], the research literature contains little discussion of systems (or system extensions) that support ancillary studies.

…We have previously reviewed open source systems for managing and integrating clinical and experimental data types; however, none of these platforms addresses the particular needs of ancillary studies[32].

…However, proprietary systems have a number of limitations, including vendor lock-in, lack of transparency, limited extensibility, and cost, as we have reviewed previously [32].

4. Finally shouldn’t the web references given eg [2-9] have accession data available (ie dates accessed) and be cited?

REPLY: Agreed, this would be nice to have. However, the journal’s required format for references does not include access dates for web references (URLs). The following link shows the format required for URL references: [http://www.biomedcentral.com/bmcmedinformdecismak/authors/instructions/researcharticle#formatting-references]

Link / URL
The Mouse Tumor Biology Database

5. On a slightly different tack I’d be interested to see the authors present some supporting evidence for the 1st sentence of the 3rd paragraph of the Background where they state “Given the importance, cost effectiveness, and prevalence of ancillary studies...”. Perhaps one simple way to do this, specifically for getting at the prevalence, is to do a Pubmed search on “ancillary studies” and state the number of hits or if there is some more specific literature out there that has reviewed this then to cite that.

REPLY: Sorry, this could have been confusing given the separation between the evidence and the claim. To beef up the claims of prevalence and importance, we added several sentences, as underlined below.
**1** Ancillary studies allow researchers to leverage the high-value data collected as part of a primary clinical trial or observational study, augment this data with additional measurements, and answer questions that were not part of the primary study design[1]. For example, an ancillary study might identify the characteristics of individual immune responses or viral types that contribute to vaccine failure or success; elucidate mechanisms of treatment response; or identify biomarkers associated with positive outcomes. Results from such studies can advance translational research and point the way to better treatments and trials. Ancillary studies can provide these benefits in a cost-effective manner because the bulk of study data has already been collected and, typically, the primary study has already shown a result worthy of further investigation.

**2** The NIH considers studies that reuse clinical trial or observational study data to be sufficiently desirable and common that it provides numerous, ongoing grant mechanisms[2–9] for funding them. Searching PubMed for the term “ancillary study” and its plural produces 892 results ([10], October 2012). A recent review of the emerging field of clinical research informatics identifies secondary data use as a key pain point[11, 12]. From a broader perspective, reuse of scientific data has become a priority and a concern across fields of all science, not just biomedical research studies[13, 14].

**3** For primary studies and trials, clinical trial management systems (CTMSs) are widely used to enhance efficiency, reduce costs, comply with regulations, and speed up data analysis. Today, researchers have a wide range of CTMS options at their disposal, from proprietary solutions such as Oracle Corporation’s Oracle Clinical[15] and Phase Forward Clintrial[16] to open source solutions such as TrialDB[17] and OpenClinica[18, 19]. CTMSs are actively studied and developed by the academic community[20–24].

**4** In contrast, with the exception of our own conference abstract[25], the research literature contains little discussion of systems (or system extensions) that specifically address the cradle-to-grave needs of ancillary studies. The open source i2b2 system[26–28] is particularly noteworthy for its support for de-identified cohort discovery across federated patient information repositories. However, descriptions of i2b2 focus primarily on the repurposing of information and material by-products of health care delivery[26–28], not on the full set of scenarios surrounding reuse of the products of trials and studies.

**5** **Given the importance, cost-effectiveness and prevalence of ancillary studies**, as well as the wide use of CTMSs, it is surprising that general-purpose
ancillary study management systems (ASMSs) have not been discussed in the research literature, either as stand-alone systems or as extensions to CTMSs. Just like CTMSs, ASMSs have the potential to ease the cost, administrative burden, and expertise required to execute ancillary studies. Given this potential, here we review the core scenarios that a general-purpose ASMS should support to provide the greatest benefits. We focus specifically on differentiating requirements for managing ancillary study data (vs. primary study data) to illustrate how the requirements for an ASMS extend beyond those of primary study management systems.

We also used your nice suggestion to quote PubMed search figures in another section. The “Naming considerations” section of the Discussion section now includes the following:

The appearance of the term ancillary studies in NIH grant announcement titles[2–9] and the frequency of this term in PubMed searches[10] (892 hits, as of October 2012) confirmed that it is in common use.

6. Very minor point when they mention stats packages they cite something called SPS whereas I’m certain they meant SPSS.

REPLY: Thanks for catching this typo. Fixed.

7. With respect to the title I believe that whilst the authors allude to Options to a limited extent the plural is not fully justified and maybe a more apt title might be “Ancillary Study Management Systems: Review of Needs and Possible Software Solution”

REPLY: The title has been cut down to:

Ancillary Study Management Systems: A Review of Needs

This change de-emphasizes the ASMS under development, as desired by Reviewer #3.