

# **Sophie 2.0: From Projects to Publishing Initiative Two, Part One**

## **Marketing Analysis Results**

**October 1, 2011**

A report of results in fulfillment of a grant to the University of Southern California  
from The Andrew W. Mellon Foundation

# Table of Contents

- Introduction ..... 3**
- Publisher Feature Survey Results ..... 4**
  - Introduction .....4
  - The Design of the Survey .....4
  - Survey Questions Listed by Highest Agreement .....5
  - Significance of the results .....10
  - Verbatim Open-Ended Responses to Survey.....11
  - Statistical Analysis Calculation Summary.....12
- Prioritized List of Sophie Feature Candidates ..... 14**
- Market Research Analysis Report..... 17**
- Intellectual Property ..... 20**
- Initiative Two Staff and Contributors ..... 20**
- References ..... 21**

# 1. Introduction

The Andrew W. Mellon Foundation funded two initiatives regarding the Sophie ebook authoring and publishing system. Further, the second initiative was divided into two parts. This document reports results from the second initiative, part one, that surveyed the publishing community to determine new feature needs and methods for expanding the Sophie user base. Part two of the initiative is reported in a separate document (see Reference 1).

The second initiative, part one (Marketing, Promotion, and Dissemination to Publishers), was originally a one-year project that began June 1, 2010, and was extended at no additional cost through September 1, 2011. Initiative Two is divided into two parts, the originally proposed research in Marketing, Promotion, and Dissemination to Publishers, and an additional Platform Survey of iPad and Android devices as platforms for Sophie-generated books.

The Marketing part of the initiative encompassed both market research to determine enhancements to Sophie to increase its value to publishers and market-based efforts to expand knowledge of the benefits of Sophie, while simultaneously acquiring market information, particularly in publishing (see Marketing Analysis Results below). The research was initiated to determine publisher attitudes and needs concerning new publishing tools, such as Sophie, and to determine the market impact of adding specific features to Sophie to guide future selection of features to for development.

Marketing research involved finding companies and individuals in decision-making roles in electronic publishing, primarily through Internet search and previous contacts, investigating how their businesses could benefit from Sophie, especially with a new publishing orientation. Data from this research and personal contacts known to the researchers informed the marketing results shown in the Sophie Market Analysis Report (see below).

However, most contacts, while interested in the subject and willing to provide their expertise, were not in a position to discuss issues that had a publishing business component, and none were willing to speak for attribution.

It was decided, in order to gain the information proposed for the grant, to augment the originally planned effort with an Web-based survey covering the major evolving needs of the publishing community in addition to the originally planned market research, which was continued, primarily by contacts, Internet search and telephone, which yielded useful information. A survey not planned for in the grant proposal, and herein referred to as Publisher Feature Survey, was designed and completed to supplement the lower level of contact input expected. This survey addressed the evolving needs of the publishing community statistically. It was designed, made available to qualified marketing professionals, and analyzed during the grant period. The survey comprised 20 specialized questions designed to determine attitudes that would help direct the future development of Sophie and determine the best ways of marketing it. The survey was

sent to 3500 individuals on the mailing list of a publishing industry e-magazine daily, *Publishing Perspectives*, arranged by Nick Matelan. The results of the survey (see Reference 1, the Publisher Feature Survey section) were statistically significant, giving insight into the needs of publishers, and provided additional guidance about the marketing strategies for Sophie most likely to produce the best results. The results added to the information used in prioritizing future work on Sophie features (see Reference 1, Prioritized List of Sophie Feature Candidates).

The research and survey information, using these multiple means of collection, provided information of higher reliability than using only one method or the other. Design of the survey, hosting it on the Web, gaining access to the mailing list, and processing the results were accomplished at no additional cost.

## **2. Publisher Feature Survey Results**

### **2.1. Introduction**

Discussions of perceived needs were conducted with contacts in the publishing sector, and results on these conversations and Internet research were used to select the questions in this Survey.

In addition to supplementing knowledge of the publishing market regarding electronic tools, one of the goals was to gauge the perceived needs of professional publishers for features and capabilities useful in publishing electronic books. The survey was stated in general terms, without direct reference to Sophie, in order to obtain an unbiased opinion. Responses were then extrapolated to specific Sophie features that could satisfy needs (see the Prioritized List below).

A sample of international publishing professionals from trade, university, and academic publishing took the survey, providing an overview of the importance of each tested feature. A link to the survey was sent out to a select list of publishers and decision-makers at publishing companies via email between June 20 and July 14, 2011.

The publishing e-magazine *Publishing Perspectives* sent email invitations, hosted the test Website, and provided raw survey results to Dr. Matelan, who compiled and analyzed the results.

The survey questions, responses, analyses, and the verbatim open-ended responder comments are included below.

### **2.2. The Design of the Survey**

The Publishing Features Survey is very different in intent and design from the first survey in this Initiative. Survey I on iPad – Android usefulness was very short and very narrowly focused. It had the single purpose of providing data for making an informed

decision about adding iPad and Android functionality to Sophie based on the importance of these platforms to decision makers in the electronic publishing Industry.

Publishing Features Survey, however, was intended to determine and rank the perceived usefulness of a relatively large number of detailed features, and to identify trends that could inform both the selection of new features for Sophie and the best initial methods for marketing Sophie to gain a larger user base.

This different survey orientation, although sent to a similar, but much larger, target publisher population, required a very different design strategy. Twenty relatively complex statements were presented. Each question required the user to assess each statement by importance in one of five categories: very important, somewhat important, no opinion, somewhat unimportant, and not at all important.

Targeting a specific population minimized coverage bias stemming from samples not representative of the population. Sources of coverage bias are typically a less carefully targeted respondent population, unreliable contact information, or lack of incentive or motivation to respond. It became clear that determining the correct non-biased discriminator for isolating decision makers was the overriding issue in crafting a meaningful survey.

USC staff on this project isolated 300 good candidates. The publishing insider blog, Publishing Perspectives, was asked by Nick Matelan to provide their extensive mailing list at no cost as additional responders for the survey. This provided a large population of 3500 thought leaders in print and electronic publishing, who actually make decisions about which books and tools to use in making decisions about implementation of new features and refinements for Sophie. Responders were all qualified professionals at all sizes of publishing organizations, with such titles as: Vice President International Sales, Publisher, Editor in Chief, Vice President Sales, and Publishing Director, and the size of the population meant that coverage was drawn from most large and mid-sized publishers.

The practices guiding survey design were:

- Find contacts that are likely to have an interest
- Clearly state the purpose
- Avoid coverage bias
- Provide a respondent-friendly survey questionnaire. The question asked must be clear, non offensive and easy to respond to for the subjects under study.

It was particularly important that the questions be as simply stated as possible, since they were in many cases complex, and in every case intended for experts in the publishing field.

### **2.3. Survey Questions Listed by Highest Agreement**

Respondents were asked to assess the importance of each of 20 propositions. A majority of respondents agreed with each of the 20 propositions. The list below, providing the statistical analysis of the results, is ranked according to the level of agreement that responses demonstrated for each proposition, with the highest level of agreement first, rather than in the order of the questions as presented in the survey.

#### Question 2. Intuitive authoring and reading interfaces

With a 95% confidence, between 89% and 95% of a population of 3,500 individuals with the same electronic publishing familiarity would respond Important or Very Important to this proposition (representing a confidence interval of 5.24 on 92.0% positive responses).

#### Question 17. Techniques to promote a consistent reading experience on multiple platforms, like multi-channel publishing to variety of devices; HTML5; readability on iPad, Android and PCs

With a 95% confidence, between 84% and 90% of a population of 3,500 individuals with the same electronic publishing familiarity would respond Important or Very Important to this proposition (representing a confidence interval of 6.78 on 94.6% positive responses).

#### Question 3. Workflow optimization features like change tracking, wizards, book templates, multiple layout options

With a 95% confidence, between 80% and 88% of a population of 3,500 individuals with the same electronic publishing familiarity would respond Important or Very Important to this proposition (representing a confidence interval of 7.08 on 84.0% positive responses).

#### Question 5. Support for ePUB 3 format

With a 95% confidence, between 78% and 86% of a population of 3,500 individuals with the same electronic publishing familiarity would respond Important or Very Important to this proposition (representing a confidence interval of 7.42 on 82.0% positive responses).

#### Question 9. Automatic index, glossary, table of contents and page numbers; spell checker; internal links; copyright notices

With a 95% confidence, between 78% and 86% of a population of 3,500 individuals with the same electronic publishing familiarity would respond Important or Very Important to this proposition (representing a confidence interval of 7.58 on 85.4% positive responses).

Question 13. Usability and accessibility, including automatic software updates; accessibility tools; adaptive screen sizing; customizable shortcuts

With a 95% confidence, between 77% and 85% of a population of 3,500 individuals with the same electronic publishing familiarity would respond Important or Very Important to this proposition (representing a confidence interval of 7.78 on 85.3% positive responses).

Question 10. Text structuring features like numbered lists, bullet lists, tables, and hyphenation

With a 95% confidence, between 75% and 83% of a population of 3,500 individuals with the same electronic publishing familiarity would respond Important or Very Important to this proposition (representing a confidence interval of 8.04 on 82.3% positive responses).

Question 7. Protection of published content, including DRM support, permission settings for readers and authors, access restrictions on specific sections or chapters

With a 95% confidence, between 73% and 81% of a population of 3,500 individuals with the same electronic publishing familiarity would respond Important or Very Important to this proposition (representing a confidence interval of 8.13 on 77.0% positive responses).

Question 1. Open systems (rather than proprietary systems), open book formats, and software

With a 95% confidence, between 70% and 79% of a population of 3,500 individuals with the same electronic publishing familiarity would respond Important or Very Important to this proposition (representing a confidence interval of 8.37 on 75.0% positive responses).

Question 15. Facilitated integration into existing publishing workflow, including content debugging, integration with CMS platforms, peer review systems, self-publishing

With a 95% confidence, between 70% and 78% of a population of 3,500 individuals with the same electronic publishing familiarity would respond Important or Very Important to this proposition (representing a confidence interval of 8.70 on 77.9% positive responses).

Question 6. Social reading features including sharing on social networks and in-book commenting

With a 95% confidence, between 68% and 76% of a population of 3,500 individuals with the same electronic publishing familiarity would respond Important or Very Important to this proposition (representing a confidence interval of 8.67 on 72.0% positive responses).

Question 16. Integrated access to external content collections and libraries like reader's library, search in library catalogues, similarity suggestions, tagging, and rating

With a 95% confidence, between 67% and 76% of a population of 3,500 individuals with the same electronic publishing familiarity would respond Important or Very Important to this proposition (representing a confidence interval of 9.00 on 74.7% positive responses).

Question 12. Reader notifications of content updates and releases of periodicals, including emails and RSS feeds for book updates; subscription for periodicals

With a 95% confidence, between 65% and 75% of a population of 3,500 individuals with the same electronic publishing familiarity would respond Important or Very Important to this proposition (representing a confidence interval of 9.04 on 72.9% positive responses).

Question 14. Incorporation of enhanced content types, like scientific, mathematical, and other discipline-related notation; Web page content; 3D displays; video capturing; video and audio streaming; animation

With a 95% confidence, between 65% and 75% of a population of 3,500 individuals with the same electronic publishing familiarity would respond Important or Very Important to this proposition (representing a confidence interval of 9.09 on 73.7% positive responses).

Question 4. Built-in media manipulation, including basic image and video editing, scripting

With a 95% confidence, between 64% and 73% of a population of 3,500 individuals with the same electronic publishing familiarity would respond Important or Very Important to this proposition (representing a confidence interval of 8.94 on 69.0% positive responses).

Question 11. Support for 3rd party software, like character recognition software, image-editing software, video editing software; video capture software, thesauri, spellcheckers

With a 95% confidence, between 61% and 71% of a population of 3,500 individuals with the same electronic publishing familiarity would respond Important or Very Important to this proposition (representing a confidence interval of 9.35 on 68.8% positive responses).

Question 18. Features to facilitate discovery of published content, like SEO and similarity suggestions

With a 95% confidence, between 61% and 71% of a population of 3,500 individuals with the same electronic publishing familiarity would respond Important or Very Important to this proposition (representing a confidence interval of 9.55 on 71.7% positive responses).

Question 8. Collaborative authoring features that allow multiple authors to work on content, including simultaneous editing, revision history, and instant messaging

With a 95% confidence, between 59% and 69% of a population of 3,500 individuals with the same electronic publishing familiarity would respond Important or Very Important to this proposition (representing a confidence interval of 9.27 on 64.0% positive responses).

Question 20. Features to support self-publishing

With a 95% confidence, between 57% and 67% of a population of 3,500 individuals with the same electronic publishing familiarity would respond Important or Very Important to this proposition (representing a confidence interval of 9.97 on 67.4% positive responses).

Question 19. Built-in translation support for multiple languages, including right-to-left languages

With a 95% confidence, between 52% and 62% of a population of 3,500 individuals with the same electronic publishing familiarity would respond Important or Very Important to

this proposition (representing a confidence interval of 9.98 on 62.0% positive responses).

## 2.4. Significance of the results

Several factors are significant in evaluating these results:

- The large skewing of the expressed opinion

The respondents agreed with a rank of important or very important for every proposition. In addition, these selections were not ambiguous, being returned at lower bound of near or greater than 70%.

- Low response rate

Response rates vary widely depending on the type of relationship the survey provider has with potential respondents. The type of relationship of the survey provider to the respondents is an indicator of the level of response to be expected in relation to other surveys of the same type, and provides a measure of response quality. This survey was of type 'no relationship'. This is the most difficult situation - obtaining responses from those with whom the survey provider has no relationship. No relationship surveys typically have a response of 0% to 20%. The population of Publishing Features Survey was of the 'no relationship' type since the members were not from any unifying organization. There was no common bond between responders other than working in the same field at an upper level. Response rates for surveys with more than 10 questions are commonly below 5%. Response rates for complex questions in such a large number of questions is lower still. Because the survey comprised 20 complex questions, the small number of returns was expected, and made getting a larger population a major goal of the project. The population was 3500 publishing professionals. The large skewing of the answers to important and very important has made the low response rate statistically significant. The return of 100 responses with the demonstrated answer skewing provides a high, 95%, confidence level that if all 3500 people had responded, the aggregate answers would be within plus or minus 2 to 5% of the returned survey responses.

- Quality of the respondent sample.

All members of the sample were of a level to speak knowledgeably about ebook platforms in qualified companies and institutions. It was a random sample within the bounds of the expert population.

- Relationship of the sample size to the population size

Research in September 2010 by the Frankfurt Book Fair involving all large and midsize publishers in the United States and the United Kingdom of senior executives indicated that the population of such individuals is approximately 1500. These two countries are among those with highest level of adoption of digital publishing products worldwide. This survey was to obtain sufficient response to represent the views of a population of 3,500 such experts with a 95% confidence.

## 2.5. Verbatim Open-Ended Responses to Survey

The following are unedited responses to the invitation at the end of Publishing Features Survey: "Any additional features you would like in an e-publishing tool not covered in this survey? Please list them here."

- Protection of author intellectual property rights and guards against theft
- Good coding interface. Make sure the user is able to see the code at all times. Validation tools are a must.
- Integration with some basic layout tools for those who can't handle the basic HTML and need embedded objects, etc. may have been included, but if not, would certainly be a nice-to-have. If you can actually accomplish the translation support, that's gold right there - perhaps even make that a separate, integratable tool?
- Make All of the Above Easy to do for a Newbie to Electronic Media.
- Ideally, we need something that has track changes and comments as in Word; that create clean XML files; can create POD files for Lulu and its ilk; can create pdfs for traditional offset printing; can create various ebook formats (with and without DRM); can handle ISBNs and create barcodes from them; and can upload all of those types of files to the relevant printers and ebook stores, along with the images and text needed to list them in the stores. The only translation support we'd need is an ability to export to Trados files, which we could handle from the
- That covers the waterfront of known desired features.
- Support for simple animation is deeply important for the children's market, which is where I am focused.
- On the retail side, small fiction publishers suffer from not being on front pages. Retailers say they post top-selling print novels on front pages. When a consumer looks for historic fiction, or any other genre, small publishers titles are listed behind large publishers' back lists. I believe we need keyword searches for fiction - rather than simply asking consumers to search through a particular genre.
- The software should be able to support Afrikaans including all the special characters that we have.
- Ability to export as pdf.
- It should be easy to use
- Make it easier and cheaper

- I think you've covered the most important from my perspective as a small, independent e-publisher of both ebooks and their coordinating print books. The absolute #1 item of priority on my list would be open-source software and compatibility across all e-reader and ebook aggregator platforms.
- Cross promotion tool for one format linking to another, adaptable content platform which allows for conversion compatibility from one OS to another.
- Metadata generation tools
- Regularly updated standardization guidelines for e-publishing across platforms (Kindle, Nook, iPad, etc.). Free online database and rating system (by authors and publishers) of companies/organizations offering ebook conversion services.
- Ability to use page layout software such as InDesign without having to convert to Word
- Copyright protection was covered, but is vitally important. Pirated content must be stopped.
- Equation support, table support, roundtripping between CMS and layout software as text edits are needed. Affordability. Tech support.
- Software that works on both Mac and PC systems, that can output into epub and mobi and any other necessary ebook format, which is very user friendly, creates table of contents (both inline and ncx). calibre is the best so far, but doesn't help with the table of contents.
- One of your questions groups 3D imaging and specialized notation (and citation?). I would argue that notation (and citation), while not yet standardized and possibly difficult to create, are not "enhanced" functions of digital content, but core to academic communication. If they had been separate from 3D images, my rating for the actual enhanced content would have been much lower in priority.

## **2.6. Statistical Analysis Calculation Summary**

The following table shows the results and calculations made on the responses returned for the iPad / Android device survey.

## Analysis Summary

Rate the following features on a scale of 1 to 5 in terms of importance (with 5 being very important and 1 being not at all important).

Answer Options	1 - Not important	2 - Somewhat unimportant	3 - No opinion	4 - Somewhat important	5 - Very important	Positive = 4 + 5	Response Count	Sum of 4 and 5	Conf Interval	Lower Bound	Upper Bound	Question #	% Positive
Open systems (rather than proprietary systems), open book formats and software	8	3	14	28	47	75	100	75	8.37	70.82	79.19	1	75.0%
Intuitive authoring and reading interfaces	1	2	5	27	65	92	100	92	5.24	89.38	94.62	2	92.0%
Workflow optimization features like change tracking, wizards, book templates, multiple layout options	1	6	9	32	52	84	100	84	7.08	80.46	87.54	3	84.0%
Built-in media manipulation, including basic image and video editing, scripting	9	10	12	34	35	69	100	69	8.94	64.53	73.47	4	69.0%
Support for ePub 3 format	1	0	17	32	50	82	100	82	7.42	78.29	85.71	5	82.0%
Social reading features including sharing on social networks and in-book commenting	5	11	12	45	27	72	100	72	8.67	67.67	76.34	6	72.0%
Protection of published content, including DRM support, permission settings for readers and authors, access restrictions on specific sections or chapters	5	10	8	25	52	77	100	77	8.13	72.94	81.07	7	77.0%
Collaborative authoring features that allow multiple authors to work on content, including simultaneous editing, revision history and instant messaging	8	15	13	35	29	64	100	64	9.27	59.37	68.64	8	64.0%
Automatic index, glossary, table of contents and page numbers; spell checker; internal links; copyright notices	3	4	7	29	53	82	96	82	7.58	78.21	85.79	9	85.4%
Text structuring features like numbered lists, bullet lists, tables, hyphenation	4	6	7	29	50	79	96	79	8.04	74.98	83.02	10	82.3%
Support for 3rd party software, like character recognition software, image editing software, video editing software; video capture software, thesauri, spellcheckers	3	13	14	38	28	66	96	66	9.35	61.33	70.68	11	68.8%
Reader notifications of content updates and releases of periodicals, including emails and RSS feeds for book updates; subscription for periodicals	2	7	17	42	28	70	96	70	9.04	65.48	74.52	12	72.9%
Usability and accessibility, including automatic software updates; accessibility tools; adaptive screen sizing; customizable shortcuts	2	3	9	30	51	81	95	81	7.78	77.11	84.89	13	85.3%
Incorporation of enhanced content types, like scientific, mathematical, and other discipline-related notation; Web page content; 3D displays; video capturing; video and audio streaming; animation	5	11	9	32	38	70	95	70	9.09	65.46	74.55	14	73.7%
Facilitated integration into existing publishing workflow, including content debugging, integration with CMS platforms, peer review systems, self-publishing	3	8	10	32	42	74	95	74	8.70	69.65	78.35	15	77.9%
Integrated access to external content collections and libraries like reader's library, search in library catalogues, similarity suggestions, tagging and rating	3	8	13	47	24	71	95	71	9.00	66.50	75.50	16	74.7%
Techniques to promote a consistent reading experience on multiple platforms, like multi-channel publishing to variety of devices; HTML5; readability on iPad, Android and PCs	1	0	4	19	68	87	92	87	6.78	83.61	90.39	17	94.6%
Features to facilitate discovery of published content, like SEO and similarity suggestions	2	3	21	27	39	66	92	66	9.55	61.23	70.78	18	71.7%
Built-in translation support for multiple languages, including right-to-left languages	6	15	14	31	26	57	92	57	9.98	52.01	61.99	19	62.0%
Features to support self-publishing	7	10	13	28	34	62	92	62	9.97	57.02	66.99	20	67.4%

### 3. Prioritized List of Sophie Feature Candidates

Based on information from personal contacts, exemplary book authors, and the highest ranking features discovered by the survey described above, the following list prioritizes new or improved features that would enhance Sophie's market standing if incorporated.

- Design and Implement an Improved User Interface and Experience

This task is based on the following survey responses and author feedback:

Question 2. Intuitive authoring and reading interfaces

With a 95% confidence, between 89% and 95% of a population of 3,500 individuals with the same electronic publishing familiarity would respond Important or Very Important to this proposition (representing a confidence interval of 5.24 on 92.0% positive responses).

Question 13. Usability and accessibility, including automatic software updates; accessibility tools; adaptive screen sizing; customizable shortcuts

With a 95% confidence, between 77% and 85% of a population of 3,500 individuals with the same electronic publishing familiarity would respond Important or Very Important to this proposition (representing a confidence interval of 7.78 on 85.3% positive responses).

Question 17. Techniques to promote a consistent reading experience on multiple platforms, like multi-channel publishing to variety of devices; HTML5; readability on iPad, Android and PCs

With a 95% confidence, between 84% and 90% of a population of 3,500 individuals with the same electronic publishing familiarity would respond Important or Very Important to this proposition (representing a confidence interval of 6.78 on 94.6% positive responses).

- Exporting Sophie Books to the ePUB3 Format

This task is based on the following survey responses and author feedback:

Question 5. Support for ePUB 3 format

With a 95% confidence, between 78% and 86% of a population of 3,500 individuals with the same electronic publishing familiarity would respond Important or Very Important to this proposition (representing a confidence interval of 7.42 on 82.0% positive responses).

Question 14. Incorporation of enhanced content types, like scientific, mathematical, and other discipline-related notation; Web page content; 3D displays; video capturing; video and audio streaming; animation

With a 95% confidence, between 65% and 75% of a population of 3,500 individuals with the same electronic publishing familiarity would respond Important or Very Important to this proposition (representing a confidence interval of 9.09 on 73.7% positive responses).

Question 20. Features to support self-publishing

With a 95% confidence, between 57% and 67% of a population of 3,500 individuals with the same electronic publishing familiarity would respond Important or Very Important to this proposition (representing a confidence interval of 9.97 on 67.4% positive responses). It is significant that a majority even of publishing professionals agreed that the future of publishing will include significant self-publishing.

- Targeted performance optimizations based on user feedback

This task is based on the following survey responses and author feedback:

Question 2, see above. Intuitive authoring and reading interfaces

- Integration of social sharing options into the Web reader

This task is based on the following survey responses and author feedback:

Question 6. Social reading features including sharing on social networks and in-book commenting

With a 95% confidence, between 68% and 76% of a population of 3,500 individuals with the same electronic publishing familiarity would respond Important or Very Important to this proposition (representing a confidence interval of 8.67 on 72.0% positive responses).

- Features for authoring - bullets, hyphenation, generated table of contents

This task is based on the following survey responses and author feedback:

Question 10. Text structuring features like numbered lists, bullet lists, tables, and hyphenation

With a 95% confidence, between 75% and 83% of a population of 3,500 individuals with the same electronic publishing familiarity would respond Important or Very Important to this proposition (representing a confidence interval of 8.04 on 82.3% positive responses).

- Direct integration of YouTube and IA video access

This task is based on the following survey responses and author feedback:

Question 16. Integrated access to external content collections and libraries like reader's library, search in library catalogues, similarity suggestions, tagging, and rating

With a 95% confidence, between 67% and 76% of a population of 3,500 individuals with the same electronic publishing familiarity would respond Important or Very Important to this proposition (representing a confidence interval of 9.00 on 74.7% positive responses).

- Native Android reader, Incremental export with preview, and embedded fonts

This task is based on the following survey responses and author feedback:

Question 17. Techniques to promote a consistent reading experience on multiple platforms, like multi-channel publishing to variety of devices; HTML5; readability on iPad, Android and PCs

With a 95% confidence, between 84% and 90% of a population of 3,500 individuals with the same electronic publishing familiarity would respond Important or Very Important to this proposition (representing a confidence interval of 6.78 on 94.6% positive responses).

- Multi-page timeline, chapter frames, a predefined sample book and page templates

This task is based on the following survey responses and author feedback:

Question 3. Workflow optimization features like change tracking, wizards, book templates, multiple layout options

With a 95% confidence, between 80% and 88% of a population of 3,500 individuals with the same electronic publishing familiarity would respond Important or Very Important to this proposition (representing a confidence interval of 7.08 on 84.0% positive responses).

## 4. Market Research Analysis Report

Data will be systematically recorded during calls and visits. A script will be developed to aid the presenter in collecting data. This report will contain the collected data and an analysis of Sophie in the publishing market. The report can be used to inform subsequent marketing and development activities. The data presented in this report will comprise the following:

All persons contacted were made aware that their comments could be included in a publically viewable report. None of the contactees were comfortable making comments in this manner, although all provided useful information to the questions below. However, data without attribution have less credibility than attributed answers. Given these responses, it was decided to construct and implement the publisher feature survey described above.

- Benefits expressed by potential users  
Open source, upcoming ability to build books for iPad and Android platforms, and the wide range of multimedia options were considered significant benefits. The concept of one platform from authoring to publishing was also well received.
- Negatives expressed by potential users  
No negatives were expressed, but this was due primarily to the low level of recognition.
- User views of Sophie compared to competition  
Those aware of Sophie, or made aware of Sophie, universally found it interesting and worth considering.
- Marketing collateral effectiveness with potential users  
Potential users in publishing, at the times of contact and survey, were in general unaware of marketing efforts for Sophie. Sophie data used in discussion were useful.
- Branding effectiveness with potential users  
Less than 10% of contactees had any awareness of Sophie.
- Packaging effectiveness with potential users  
Packaging effectiveness was not quantifiable with the low level of recognition.

- Labeling effectiveness to users  
Labeling effectiveness was not quantifiable given the low level of recognition.
  
- Publisher prospects with the highest probability of closing  
Current possibilities are better with very large distributors, such as:
  - Amazon
  - Google
 and with self-publishing houses, specifically:
  - Abika.com
  - Baen Books,
  - Seattle Book Company
  - Barnes & Noble Publ!
  - EBSCO Publishing
  - Atria Books
  - eBook Architects.
  
- University press prospects with the highest probability of closing on some level of cooperation
  - Harvard Business Review
  - Oxford University Press
  - Cambridge University Press
  - Princeton University Press
  - New York University Press
  - University of Pennsylvania Press
  - Temple University Press
  - Rutgers University Press
  
- Partnership candidates with the best fit
  - Amazon Kindle eBooks
  - Google
  - Bedford, Freeman, Worth
  - Harvard Business Review
  - Macmillan Publishers
  - Oxford University Press

Atria Books  
Barnes & Noble  
Pearson  
Sony

- Decision makers at each prospect

Although management structure for candidates was researched, the best direct contacts were not determined.

- Preferred monetization methods

The commonly discussed ebook sales business approach was the subscription-based model, although publishers contacted would say little about such business issues. Amazon appears to be pursuing this approach, for example. Although there are issues to be resolved concerning licensing and royalties, one unknown is what pricing structure would be perceived as reasonable by the public while providing a justifiable return.

- Effect of Open Source licensing on acceptance

Open Source tools were very widely thought to be better (for both business and technology reasons) than proprietary tools.

- Best potential techniques for advertising, sales promotion, and public relations

There was no clear consensus.

- Expected market segmentation and best product positioning

The market is thought, by our contactees, to be bifurcating into large on-line sources of eBooks and providers of self-publishing services. Both of these segments could benefit from Sophie for the upper tier of books that would take advantage of from advanced integration of image, sound, and video.

- Potential legal issues associated with licensing, copyright and meta-tagging

Several legal issues are unique to electronic books and publishing. The following were raised:

Content (multimedia elements as well as text) in ebooks may include links to other sites. Linking from site A to site B could involve an actual transfer to site B, or it might involve only the inclusion of content from site B “framed” by the interface of site A. Moreover, a transfer link may be to the homepage of site B, or to an internal page within site B’s set of pages (a “deep link”). Generally, linking to a homepage does not pose a problem. Some commercial sites, such as on-line newspapers or magazines, may object to “deep linking,” but there is no settled U.S. law regarding its illegality, and other nations have come to disparate conclusions about the legality of the practice. The display of content from another site on one’s own site (“framing”) without license to do so is likely to be considered a violation of copyright law, although litigation has not produced a

definite decision in the U.S. (*Futuredontics Inc. v. Applied Anagramic Inc.*, 45 U.S.P.Q. 2d 2005 (C.D. Cal. 1998)). The best policy regarding any use of the content of another is to obtain permission from the rights holder, in the form of a linking or licensing agreement.

Content in ebooks may also include meta-tags that, while not displayed to the normal viewer, are readable by software such as search engines. Meta-tags may contain content such as lists of keywords, for example, to increase the frequency with which the documents are identified by various search terms, although some search engines no longer include results from meta-tag content. There have been attempts to copyright lists of keywords, but no such claim has been successfully asserted.

Another issue arising from the use of meta-tags is the inclusion of trademarked terms as keywords. Such inclusion can be considered a violation of the trademark owner's rights, if it has the effect of confusing the consumer (infringement) or if the use is likely to dilute the distinctiveness of a famous mark (dilution).

One particular defined meta-tag is the COPYRIGHT meta-tag, which may contain a copyright declaration with, for example, the associated rights owner's name. The COPYRIGHT meta-tag is not a substitute for a viewable Copyright notice. However, copyright law does not require such a notice, or any explicit assertion of protection. Every creative work published in the U.S. since 2002 is the copyright of its author, by default, for the life of the author plus seventy years.

## **5. Intellectual Property**

Sophie 2.0 is free and open source; the code and documentation can be found within the "Developers" section on the Sophie 2.0 website: [sophiecommons.org](http://sophiecommons.org). The code is licensed under the ECL 2.0, and all other information is licensed under the Creative Commons By Attribution 3.0 United States License. Licensing information is available from the Sophie Website.

## **6. Initiative Two Staff and Contributors**

Elizabeth Daley and Holly Willis, Co-Principal Investigators  
Nick Matelan, Marketing research and promotion director  
David Weitzner, Marketing advisor  
Laura Cechanowicz, USC Graduate Student, Survey Research  
Hannah Johnson, Publisher, Publishing Perspectives  
Thomas Minkus, Vice President, Frankfurt Book Fair  
Ludmil Pandeff and the Sophie development group at Astea Solutions

## 7. References

Reference 1: Initiative Two, Part Two, Platform Survey Results