Internet Speculative Fiction
DataBase

ISFDB2 Concept of Operations Document

REVISION HISTORY

Version	Date	Name	Reason
1.0	07/13/2001	Al von Ruff	Initial Draft

1. Scope

1.1 Identification

This is the concept of operations (ConOps) document for the ISFDB2.

1.2 Document Overview

This document identifies the high-level needs and expectations for the ISFDB2; in particular:

- The major types of data which need to be represented in order to construct bibliographic content lists, as well as detailed author bibliographies.

- The data organization required to support an open data project.

- The methodology needed for data change requests to be implemented

- The relationship between the ISFDB2 and ISFDB1, as well as CD-ROM versions of either database.

By providing an overview of the system from a user's perspective, the ISFDB team can better understand the requirements for this database.

The primary document audience is ISFDB2 users and developers. Users of ISFDB2 will be interested in verifying that this document accurately describes their needs. Developers of ISFDB2 will use this document to drive system-level requirements for the data, support tools, and the site layout.

An overview of the situation describing ISFDB1 and justifications for ISFDB2 are provided as background information. Following this is a high-level description of the ISFDB. The various user classes of the ISFDB are then listed, providing a perspective for each class and scenarios detailing how each class uses the system. Finally, the impacts, advantages, and disadvantages of the ISFDB are summarized.

1.3 System Overview

The overall purpose of the ISFDB is to serve as a repository for speculative fiction bibliographic data, and to provide a user interface to respond to user queries. Located worldwide, ISFDB users are interested in viewing the following types of information:

Author Bibliographies - An author bibliography is an annotated listing of the works produced by an author. Annotations may include the publication history of the work, notes and synopsis, a list of awards won by the work, and the location of reviews concerning a particular work. In addition to bibliographic data, the bibliography may also contain personal details about the author such as birthdate, birthplace, and a condensed biography.

Content listings - A content listing displays publication metadata (publisher, ISBN, page count, etc...), and a detailed list of the works contained within the publication. The works may be annotated with details such as original copyright date, location of first publication, translators, variant titles, or pseudonym resolution.

Yearly Awards Listings - While individual award annotations can be found in the author bibliographies, it is often desirable to view all of the nominations and winners in all categories of a particular award for a particular year.

Magazine Checklists - While book publications are organized around the authors and editors that create then, it is desireable to organize magazine publications with checklists. A checklist show all issues for a particular magazine, structured by year in the order of publication.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Forthcoming Books - Listing books which will be published in the coming months allows users to track the pubication status of books they are interested in.

In support of the above listed view types, the ISFDB supports the ability to search the database by author, title, or year, and provides an interface to submit or correct data.

2. References

1. *IEEE Standard 1362-1998: IEEE Guide for Information Technology -- System Definition -- Concept of Operations (ConOps) Document*, The Institute of Electrical and Electronic Engineers, Inc.

2. John Wenn. Various speculative bibliographies from the early to mid 1990's are archived at **http://www2.lysator.liu.se/sf_archive/sub/jwenn.html**

3. Saul Jaffe. *SF-Lovers Digest*. Back issues are archived at **http://sflovers.rutgers.edu/**

4. John Leavitt. *Speculative Fiction Clearing House*. This site has been re-established as Steampunk at **http://www.steampunk.com**

5. Ahasuerus, aka *The Wandering Jew*, was a high-profile net personality who cruised **rec.arts.sf.fandom** and **rec.arts.sf.written** during the time the ISFDB was created. Since he didn't actually write any code for the ISFDB, Ahasuerus continues to downplay his role in its creation. He nonetheless established the bulk of the system requirements for the ISFDB.

6. *The Internet Movie Database*, **http://www.imdb.com**

7. Stephen T. Miller, William G. Conento. *Science Fiction, Fantasy, & Weird Fiction Magazine Index: 1890-1998*. Locus Press (P.O. Box 13305 - Oakland, CA 96601), July 28 1999.

8. *ISO 4217. Codes for the representation of currencies and funds.*

9. *ISO 639. Codes for the representation of names of languages.*

3. Current System or Situation

3.1 Background, Objectives, and Scope

While the discipline of creating bibliographies and indices for the speculative fiction genre has always been rather mature, there was little such material available on-line by the mid 1990s. The best on-line bibliographic resource at the time was the series of bibliographies created by John Wenn [2], which were posted periodically to the SF-Lovers digest [3] in the late 80's and early 90's.

In early 1994, Al von Ruff created a private searchable database of select genre-related awards information. By September of that year, the database had been expanded to include most genre-related winners and nominees with data supplied by David G. Grubbs. By the end of 1994, a web version of these private tools was created. These tools were the precursor to the ISFDB.

In 1995, one of the most popular SF portals was *The Speculative Fiction Clearing House*, edited by John R. R. Leavitt [4]. The site had sections devoted to theme bibliographies, conventions, bookstores, writing advice, and awards. In the spring of 1995 the *Clearing House* was looking for a set of awards database tools, and the pre-ISFDB tools were offered for use there. The site however, was looking for a more tightly integrated database - the *Clearing House* was beginning to put magazine content listings online (via manually constructed webpages), and wanted to link those in some way with the awards information. The pre-ISFDB tools were never utilized by the *Clearing House*.

In the summer of 1995, Al von Ruff and Ahasuerus [5], armed with lessons learned from the *Clearing House*, incorporating John Wenn's bibliography layout, and inspired by the look-and-feel of the *Internet Movie Database* (IMDB) [6], began to design and build the ISFDB (*Internet Speculative Fiction Database*). The ISFDB 0.1 went online 8 Sept 1995, and went through numerous revisions prior to its public launch in Jan 1996.

Starting as a home page at a local ISP in Champaign Illinois, the design of the ISFDB was impacted by various deficiencies commonly found in ISPs at the time, including: limited memory, lack of real database support, disk space limitations, and limited processor resources. These limitations fundamentally shaped the architecture and operation of the ISFDB.

3.2 Operational Policies and Constraints

The database is maintained and updated by a single person; this person has editorial control of the ISFDB. Diskspace, monthly throughput rates, and processing resources are constrained as per agreement with the ISP hosting the ISFDB.

3.3 Description of the Current System or Situation

Section 1.3 described the different types of bibliographic data that users of the system are concerned with. The data is displayed to the user in the form of HTML-based web pages. These web pages are dynamically constructed upon request by the user from data stored in the ISFDB database.

3.3.1 Database File Structure

While commercial websites were utilizing Oracle or Informix database servers (mySQL was in its infancy it the time the ISFDB was created), these types of databases were not available at the hosting ISP. Data was instead stored in flat ASCII files (one for each table), using pipe-delimited fields. This technique was a common method of storing data on UNIX systems, and felt natural for use in this setting as well.

3.3.2 Indices

To speed lookups in the flat database files, a series of indices are utilized. Each index uses a particular table field as its key, and associates it with a series of table offsets to the appropriate records. Executables charged with creating the dynamic webpages search for the key in one or more of these indices, then quickly seek to the needed records in the target tables.

3.3.3 Database Tables

The ISFDB uses six main database tables to respond to queries: Authors, Titles, Publications, Awards, Reviews, and Interviews. Figure 1 shows how different web pages can be constructed by selecting the required records from the various tables.

3.3.4 Table Descriptions and Rationale

A description of the six tables follows, along with the rationale for it:

- **Authors** - The Authors table contains information about the author, such as birthdate. This information is persistent across all of the publications associated with the author. The primary key for this table is the author's canonical name. This is the name that the author most commonly publishes under, not necessarily the authors legal name. The canonical name may be a pseudonym.

- **Publications** - The Publications table contains all of the information about a particular publication, aside from content. This includes information such as publisher, page count, cover artists, ISBN, and so on.

- **Titles** - The Titles table contains a normalized list of works. For instance, a particular story may have been published in numerous locations. In order to save space, instead of listing the title and author under each publication it was published under, it instead takes up a single record in the Titles table. In order to link the title with a particular publication, each title record contains a field which points to the publications it was published under. Every publication has a unique identifier which the title can refer to.

- **Reviews** - The Reviews table is a specialized Title table. While a record in the Titles table has title and author fields, a review has an additional reviewer field. An extra field could have been added to the Titles table, but since there are very few reviews in compa-
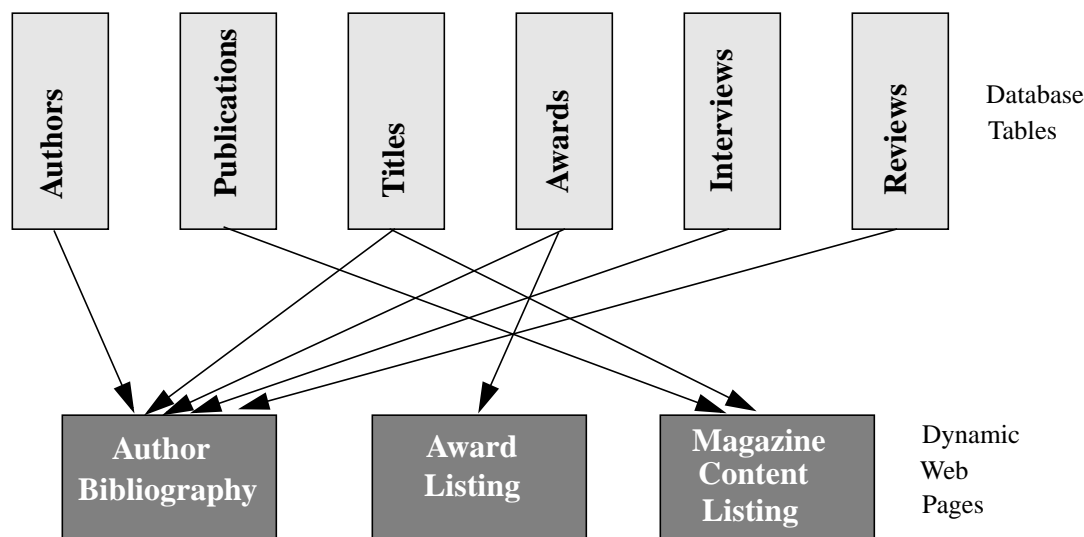

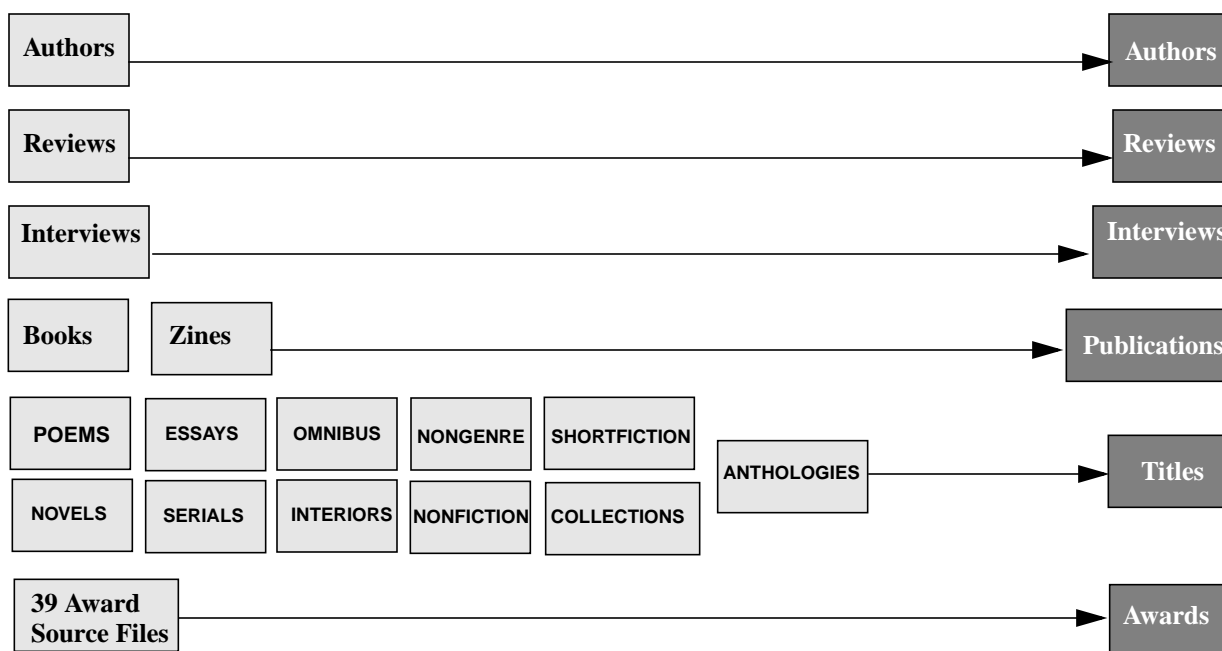**Figure 1**: Creating Specific Dynamic Pages From the Six Database Tables

Figure 2: Creating the Six Database Tables From 55 Offline Source Files

rision to the thousands of titles, it saved more space to break out the reviews into its own table.

- **Interviews** - Like the Reviews table, this is also a specialized Titles table. An interview not only has a title, it also has a subject and an interviewer to track. Same rationale as that for reviews.

- **Awards** - The Awards table contains a list of award records. Each record has an associated type (Hugo, Nebula, etc), as well as the awarded title, author, and what level of award was won.

Although the format of the six database files have good performance in the online environment, they aren't very good documents for humans to edit. Many of the pipe-delimited fields can be empty - if a person wants to add data to a particular field of a particular record, they may have to count pipes to find the field. This tends to be error-prone, so another version of the files exist in an off-line format, which is suitable for editing. Furthermore, the larger database tables are broken down by types into smaller files, making them easier to edit on systems with small memories. The very first revision of the ISFDB had only the pipe-delimited files, and it was due to the great difficulties encountered while editing those files that the off-line format was created.

The six online database tables then are created from 55 offline text files by "compiling" them with a set of tools constructed specifically for this pupose. The relationship between the offline text files and the online database files is shown in Figure 2. Note that although the offline files are much smaller than the aggregate online files, they can still be rather large. The SHORTFICTION file is nearly 5 Megabytes in size, with about 53,000 records. This large size can make the data awkward to pass about in email, and can bog down (or even crash) certain text editors.

Another down side to slicing up the tables in this manner is the number of files which need to be touched in order to enter the data associated with a publication. For instance:

- **Novel** - Entering data for a novel could require touching *NOVELS* and *BOOKS*.

- **Anthologies/Collections** - Entering data for a short story collection or anthology could require touching *BOOKS, COLLECTIONS, SHORTFICTION,ESSAYS*, and *POEMS*.

- **Magazines** - Entering data for a magazine could require touching *ZINES, SHORTFICTION, ESSAYS, REVIEWS, INTERVIEWS, INTERIORS, POEMS,* and *SERIALS*.

Thus, adding certain types of publication data may require access to numerous database files simultaneously.

3.4 Modes of Operation for the Current System or Situation

The ISFDB is split into two distinct systems:

- **Off-line Database** - The offline database occupies about about 22 Megabytes of space (which gzips down to about 5 Mbytes). The offline database is currently located in the residence of the ISFDB editor. There are 3 main reasons for this:

  - There isn't enough space at the supporting ISP for both the on-line and off-line versions of the database.

  - The original ISP didn't have enough system memory to actually perform the building of the on-line database.

  - The merging work is conducted by the ISFDB editor, who lives in a rural area with no access to broadband services. As such connectivity is limited to short sessions of low-bandwidth dial-up service.

- **On-line Database** - The online database currently occupies about 20 Megabytes of space (which gzips down to about 7 Mbytes. Note that the compression is less efficient due to the presence of the indices).

The on-line version of the ISFDB has two distinct modes of operation:

- **Search and Display** - Users have the ability to search by title, author, year, or series. Awards for a specific year can be observed. An alphabetized author directory exists which the user can browse through. Select magazines have checklist tables which the user can browse through.

- **Data Submissions** - In data submission mode, forms are available to add/correct authors, publications, and title information.

3.5 User Classes and Other Involved Personnel

There are six broad user classes of personnel who need access to the ISFDB filesets, source code, or the web site. These are:

1. **ISFDB maintainers -** There are three primary maintainance roles for the ISFDB (note that multiple people may fulfill a particular role, or that a single person may take on multiple roles):

- The website editor, who performs database merges, integrates user-submitted data, adds new features, and performs periodic builds.

- The awards editor, who maintains an awards database seperate from the ISFDB. Awards informations is transformed into ISFDB format by David, who sends the data to Al, who integrates the data into the ISFDB.

- The forthcoming books editor, who maintains forthcoming book information in a separate location (at present in an Access database). The data is exported and translated into ISFDB format, and integrated by the website editor.

2. **Authors** - Numerous authors desire to utilize the ISFDB as a permanent bibliography resource which represents their professional output. This often includes works which are not of a speculative nature, nor associative.

3. **Fen** - Members of fandom have an interest in promoting particular authors or books, and take the time to update those bibliographic items.

4. **Bibliographers** - Numerous single-author bibliographers create detailed bibliographies, utilizing a number of online resources, including the ISFDB.

5. **Readers** - Casual readers of speculative fiction refer to the ISFDB to locate books and stories by favorite authors, to locate additional books in a particular series, to locate where a particular short story may be found, and to access the various reading lists which are generated from ISFDB data.

6. **Booksellers** - Book sellers, particularly those who sell used books, utilize the ISFDB to locate titles, and to attempt to determine the work's resale value.

7. **Bibliographic Linkers** - Numerous websites link to the ISFDB bibliographies. For instance, when con coordinators discuss writers in attendance to a particular con, they often provide a lilnk to the author's ISFDB bibliography. SFsite utilizes ISFDB bibliographies in their book reviews.

3.5.1 Organizational Structure

There is no organizational structure to the users of the ISFDB.

3.5.2 Profiles of User Classes

ISFDB maintainers - Although the ISFDB maintainers spend most of their time in the off-line version of the ISFDB, they also utilize the online version in the course of their work. When resolving pseudonyms, or determining an author's canonical name, author searches are often conducting. When pulling together variant titles for a particular work, title searches may be conducted.

Authors/Bibliographers - Authors are interested in finding, or establishing, their presense in the ISFDB. Once located, authors begin to inhance the accuracy of data already present in the ISFDB, and then expand it to be more inclusive. The degree of cooperation amongst authors varies greatly - some are interested in entering complete publications that their works appear in, while others only wish to enter data that pertains them. Authors require the ability to search for their bibliography, and the ability to input data into the ISFDB. Bibliographers need the ability to display the summary bibliography for a particular author, and from there display detailed publication information. Some bibliographers will place missing information into the ISFDB. As such, the online requirements for bibliographers is very similar to those of an individual author.

Fen/Readers - The interests of readers tend to be broader than those of an author or bibliographer. Aside from interest in multiple authors, readers also have interests in print series (which may span across multiple authors), and subject bibliographies (such as books about nanotechnology, cyberspace, or giant robots). Some readers, acting as oracles for forums (such as the USENET newsgroup rec.arts.sf.written), utilize the ISFDB to locate a title from vague clues given by a less knowledgable reader. These readers need the ability to list the stories found in a particular magazine, collection, or anthology; utilize story synopsis; and would like to access, or search for, subject or keyword information.

Bibliographic Linkers - User who wish to link to ISFDB bibliographies require the ability to search for a particular author's bibliography. Once found, this user class requires the ability to link to a persistant web page. Other bibliographic services construct static web pages from database records, meaning that a particular author's works may migrate from one static page to another across builds, making it difficult to link to that author's works.

Booksellers - Booksellers and collectors desire the ability to download descriptive information from the ISFDB via ISBN, catalog number, bar code reader, or by simple author/title search. The data would then reside in an offline (and perhaps smaller) version of the data-

base. Given a specific book, the application would have the ability to automatically look up summarize current internet offerings, information from price guides, bibliographic information, or library holdings. The application should also be able to upload new data to the online database. Note that none of the functionality is currently available in the ISFDB; booksellers are required to perform all of the above steps themselves, manually.

3.5.3 Interactions Among User Classes

As no forums for communiation are present at the ISFDB, the only conduit for interaction among user classes is in the data deposited and integrated. Interaction may occur in other venues (such as USENET newsgroups), but these interactions are not controlled or monitored by ISFDB personnel.

3.5.4 Other Involved Personnel

None.

3.6 Support Environment

Support issues, such as maintaining backups, is left solely to the hosting ISP. Between releases, backups of the ISFDB source code and data are made to Zip disks, which are stored at different physical locations (which assures at least one good copy in the event of catestrophic property loss). Upon release to the web site, an archive of the ISFDB source code and data are made to CD-ROM.

4. Justification For and Nature of Changes

4.1 Justification of Changes

Over the course of the last six years, numerous changes have been made to the ISFDB to improve its utility. There are, nonetheless, fundamental problems with the current implementation which call out for changes:

4.1.1 Need For a True Open Data Project

As an open data project, the ISFDB has not been entirely successful. While users have the ability to input data, that data is filtered through an editor before integration into the database. More importantly, the integrated ISFDB data is not readily available to the public at large. One reason for this is the lack of online space, but the real gating problem is the fact that the data lives in the off-line database, and that off-line database is not accessable from the Net. Access to the system which hosts the offline database is limited to removable media, and its bandwidth to the Internet is limited to modem speeds (broadband is unavailable in the system's area). This means that uploading the data to an accessible location could take several hours of dial-up time.

Another potential problem area is complete loss of the ISFDB data. Since there is a single data custodian, and the data is generally not available online, the data custodian becomes a single point of failure. Armed with access to the online database, a knowledgable person could reconstruct the offline database. However, since the data custodian is also the web site editor, and controls access to the website, such knowledgable people may not be able to gain access to the online database.

4.1.2 Need For an Open Source Project

Since the ISFDB is a closed system, there has been no opportunity for individuals to contribute new and interesting applications to the project. This inhibits innovation, as ideas for new features must be filtered through a single individual for implementation. A need exists to open both the online and offline toolsets to those that wish to contribute.

4.1.3 Need For an Off-Line Version of the Database

Entering bibliographic data is a time-intensive process. As many users do not yet have broadband service, extended periods of data entry can lead to large dial-up costs. In addition, some users wish to enter data from locations that have the primary sources, but do not necessarily have direct access to the online ISFDB (say from a library). These users would prefer to install the ISFDB in some form onto a laptop, make changes to the database while in the field, and then upload those changes for integration into the ISFDB.

Booksellers are looking for an extensible ISFDB - a version in which additional fields can be added to support data (such as current or estimated value of a book) relevant to booksellers..

4.1.4 Need For Instantaneous Data Submission Feedback

Since modifictions can't be made directly to the online database, and must be filtered through the editor, it can take weeks (and in some cases months - okay, in some cases *years*) for data to appear online. This not only leads to frustration on the part of the user, but can also result in multiple entries from different users for the same work. It makes it difficult for users who are trying to complete some aspect of a bibliography - as many weeks later they are unsure if they forgot to enter that data, or whether the data was lost, or if it simply hasn't been integrated yet.

4.1.5 Need to Reduce Submission Errors

The current data entry pages of the ISFDB allow the users far too much latitude when entering data, resulting in a tremendous amount of erroneous data being submitted. Each of these errors require additional bibliographic research on the part of the editor, resulting in additional delays in integrating the data. The data and its associated toolset need to be structured in such a way that minimizes errors, performs detailed error analysis, and provides error feedback to the submitter.

4.1.6 Need to Reduce Author Centricity

The current ISFDB focus on author bibliographies gives the erroneous impression that data is also structured around author bibliographies. Many users assume that data can simply be added to an author's bibliography, when in fact, the author bibliographies are dynamically constructed at runtime from publication data. These users don't understand that the way in which to extend an author's bibliography is to enter the publications in which the author's works appeared in.

This attitude has three undesireable side effects. First, it results in numerous author biographic submissions which merely consist of instructions to visit an existing web page and enter the bibliographic data found there into the ISFDB. As no staff exists to take these bibliographic joyrides, these requests are typically not granted.

The second side effect is that a great percentage of data submissions come with no publication information at all. That is, a submission for a short story is given with the title author, and date of publication - but no details on where the story was published. This results in weak bibliographies that are of little use to readers.

The third side effect of this attitude is that it doesn't help fill out the bibliographies of any other authors. Authors typically receive a contributor's copy of a publication (in some cases, receipt of the publication is the only payment the author will receive). If an author has published a story in a little-known publication, and that publication has not been cataloged due to its scarcity, then one of the few avenues available to cataloging that publication is for the author to enter the complete publication data - not just the data that pertains to the author.

4.2 Description of Desired Changes

4.2.1 Organization

An organization should be established. This organization would have legal ownership of the ISFDB data and source, would control access and licenses to the data and source, and would maintain and operate the website which presents the data and source to the users. The organization, if legally incorporated, would be a not-for-profit entity.

4.2.2 Metadata Representation

The data must be represented in a format that is portable, and easily transmitted to those participants that wish to use or modify the data. It should also use be in a non-proprietary format which is easily extended in the future.

Many sites, when faced with similar data representation requirements, choose XML. Once a DTD has been established, the data becomes non-proprietary and can be easily parsed with any number of XML-enabled tools. Data is in the form of ASCII (or UTF-8) files, which can be easily emailed if necessary.

Additionally, all data should be represented via the follow four file types:

- **Publication** - A publication record describes all of the metadata associated with a particular publication. A publication is typically a book or magazine, but the precise scope may be extended to cover electronic media which meets certain requirements. In order to facilitate the sharing of data, there is one and only one record per publication file. The type of data found in the publication record is restricted to that information which can be derived directly from the publication. For instance, title, author, page count, ISBN, the titles and authors of works contained in the publication, original copyright dates of the contents, location and creator of illustrations, and the page numbers that particular works appear on are all examples of metadata that can be objectively located in the publication. Examples of metadata that may not be objectively located in a publication are: variant title of a work found in the publication, whether a story was written under a pseudonymn or not, whether or not the publication or a work within it won (or was nominated for) an award, or what series the publication might be a part of. As such, this type of metadata should not be a part of the publication record. With the publication in hand, it should be trivial for an untrained reader to submit the metadata associated with the publication.

- **Author** - An author record describes all of the metadata associated with a particular author. The data found in such a record does not differ greatly from the present ISFDB author record. The primary difference is, again, there is one and only one record per author file.

- **Canonical** - While restrictions on the publication records make them unambiguous, it also makes it more difficult to create complete author bibliographies. Information about variant titles and pseduonymns need to be taken into account to construct anything but the most simplistic bibliography. A canonical record contains the canonical title and author name associated with a work, and then lists known variant titles, pseudonyms, awards information, the series a work might have been a part of, notes, synopsis, and perhaps subject/keyword data.

- **Series** - A print series spans multiple works, and may span multiple authors as well. Metadata about a print series includes notes, and subseries information. This metadata is stored in a series record. Starting with the highest level of the series heirarchy, a record lists all subseries, and supports the annotation of individual series with notes. A series record does not described wh,ich books are in a particuler series - it merely gives a series structure. The canonical record for a book will contain a pointer to the series name it is a part of (if any), pointing to a leaf in the series tree structure..

4.2.3 Change Request Data Representation

Not only should data be represented in XML, but modification requests to the data should also be represented in XML. These can be generated offline if need be, and transmitted to the web site either directly (through a protocol such as SOAP), or indirectly through an email service.

4.2.4 Web Site

A new web site should be established. The current website (http://www.sfsite.com/isfdb) would continue to exist in its current format, as it is well-established, and has many thousands of links pointing into it. A new web site would support the new data formats mentioned above, and will support the following major facilities:

- **Data Repository** - The data repository would hold all of the bibliographic data.

- **Documentation Repository** - The documentation repository would hold documentation on data formats, source code, user's guides, and project documentation such as this Concept of Operations document.

- **Source Code Repository** - A facility to version control applicable source code would be supplied, along the lines of that found at SourceForge.

- **Bibliographic Forums** - A forum should exist whereby people can discuss bibliographic problems and suggest solutions.

- **Application Forums** - A forum should exist whereby people can dicuss potential software applications and their implemenations.

- **Download Facilities** - A facility should exist that allows users to download data, documentation, and software.

- **Change Request Facilities** - A bug reporting facility should exist to report software errors, requests for extensions to document types, and to suggest enhancements or feature requests.

- **Upload Facilites** - A facility should exist whereby data entered offline is uploaded for integration into the database.

4.3 Priorities Among Changes

The priorities for changes to the current situation which are considered essential are:

1. **Establishment of Organization**. Before work can be split up among people, there must be people. The organization need not be formalized at first, but interested parties need to be brought together to get things rolling.

2. **Establishment of Website**. Once people have expressed an interest in working on the project, they must be given a place to work.

3. **Construction of Forums**. In the beginning, discussions will have to be held concerning the details of all aspects of the new project. The existence of forums on the website will facilitate these discussions.

4. **Document Type Definitions.** Agreement must be made an the various data formats that will be utilized by the ISFDB2. These formats must be defined before the repository canbe constructed.

5. **Creation of the Data Repository**. Once the data formats have been created, the current ISFDB1 data will be translated to the new formats and stored in the data repository.

6. **Creation of Change Request Facilities**. Once the data is installed in the repository, the actual work of modifying, correcting, and extending the data will require online tools to do so.

7. **Creation of Download Facilities**. In order to create ISFDB1 data, download facilites will be required to extract the data from the repository.

8. **Creation of Upload Facilites.** These facilities need to be present before people begin to start modifying data offline.

9. **Creation of Offline Tools**. Once the upload facilities are in place, people can begin the task of creating new data. This will require a set of offline tools.

4.4 Changes Considered But Not Included

Merging or moving the current ISFDB website.

4.5 Assumptions and Constraints

The following assumptions are applicable to the changes and new features identified above:

- There are enough preople interested in the ISFDB to take part in the project.

5. Concepts for the Proposed System

5.1 Background, Objectives, and Scope

Section 4 introduced the new elements of the proposed system. This section looks into those elements with more detail.

5.2 Operational Policies and Constraints

There are three major operational policies governing the proposed system:

- Creation of an organization that owns the data.
- Creation of an open source project for the supporting tools.
- Creation of a web site that serves as a data, documentation, and source repository.

The major constraint is the construction of a web site that supports the needed features at low cost.

5.3 Description of the Proposed System

The proposed system is shown in Figure 3. It is broken down into four main components: ISFDB1, ISFDB2, data converters, and the offline database.

5.3.1 ISFDB1

The term *ISFDB1* represents the current ISFDB which is located at **http://www.sfsite.com/ isfdb**. Thousands of permanent links into this site exist across the web, referring to numerous author and magazine pages. Moving the site to a new location would cause widespread disruption of these links, with little gain in return. As such, the current site will remain at its present location, with the following alterations:

1. The offline database filesets described in section 3.3 will no longer be utilized. The data repository will consist of an entirely new format, stored in a new location (which will be referred to as *ISFDB2*). To minimize changes to the ISFDB1, data from the new repository will be converted into ISFDB1 format whenever an update to the website is desired.

2. Modifications will no longer be integrated into the ISFDB filesets. All modification requests will be directed at the ISFDB2 data repository.

3. Modification requests generated at the ISFDB1 site will be in a formatted in the manner required by the new data repository at the ISFDB2.

4. The role of the ISFDB1 editor will shift dramatically. Presently, the editor takes the data submitted at the web site, modifies it to conform to data conventions, and then integrates the data into the offline database filesets. The new role of the editor will be to collect the data submissions, monitor them for accuracy, and then submit the data to the ISFDB2.

The flow of data from the ISFDB2 to the ISFDB1 and back again, can be seen at the bottom of Figure 3. As can been seen there, the major change is in the user interface, which collects user submissions and forwards them on to the ISFDB2. Input to the ISFDB1 is through a set of data files generated by a data converter. This converter takes the XML version of the data from the ISFDB2, and converts it into ISFDB1 format.
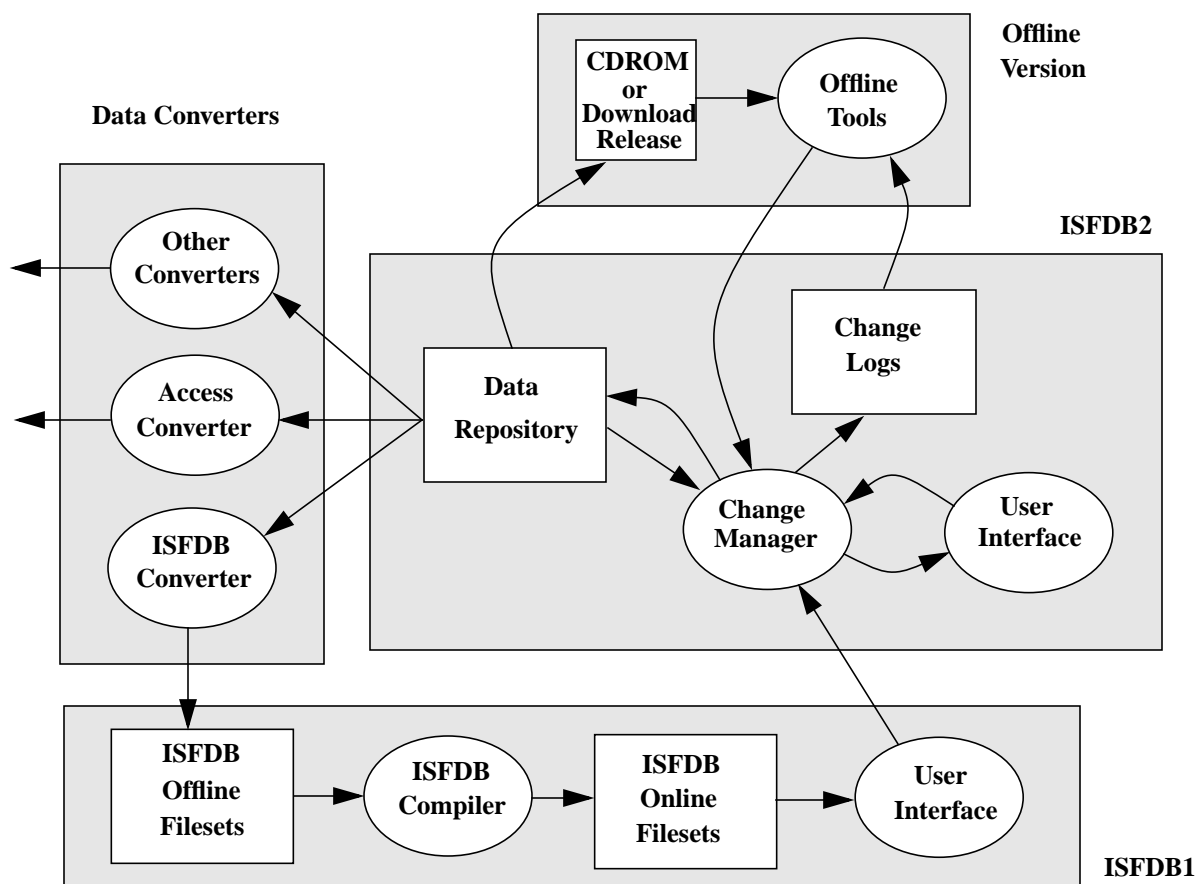
Figure 3: Block Diagram of Proposed System

5.3.2 ISFDB2

The term *ISFDB2* represents the new web site located at **http://www.isfdb.org**. The four major components of the ISFDB2 are:

- **The Data Repository** - All ISFDB data is stored in the data repository. The official repository is located online, in order to make it accessable to everyone. While the structure of the data repository is not defined by this document, it will conform to the following attributes:

  - The repository will have a fully documented and stable API.

  - Documents will enter the repository as XML-formatted data.

  - Documents will exit the repository as XML-formatted data.

- **The Change Request Manager** - The change manager is the main interface to the data repository. Utilizing the data repository API, it has the ability to extract data from the repository, and to add or modify data to the repository. All communication with the data repository will be formatted in XML. Successful modifications to the data repository will be logged.

- **Change Logs** - The change logs represent the history of changes made to the data repository. The change logs are a repository of the XML requests made to the data repository. The change logs will be used by the offline tools to create up-to-date views

of the data repository. This abililty will allow offline users to synchronize with the ISFDB2, without downloading the complete data repository, which may be quite large.

- **The User Interface** - The user interface allows registered users to create data modification requests on site, and have the requests integrated immediately

5.3.3 Data Converters

Data converters take the XML input from the ISFDB2 data repository and transform that data into a format usable by another tool. At a minimum there must be a data converter whcih targets the ISFDB1. There is already a desire for a data converter which targets Microsoft Access database format.

5.3.4 Offline Version

The offline version of the database resides on the end-uers computer system. The tools associated with the offline version allows the user to modify data without connection to the ISFDB websites. Change requests are stored locally, and uploaded to the ISFDB2 website periodically. When uploaded, these change requests are fed into the ISFDB2 change manager for integration into the data repository.

5.4 Modes of Operation for the Proposed System

5.4.1   ISFDB1

The ISFDB1 still has two major modes. These are:

- **Off-line Database** - The biggest change for the current ISFDB is that there is no longer an off-line version of the database files. Data from ISFDB2 is exported through a data converter to create the input files for the ISFDB1. These files are not to be modified - all data changes are made to the ISFDB2 filesets. Static web pages (such a magazine checklists) associated with the ISFDB1 will continue to reside in the off-line version of the ISFDB1.

- **On-line Database** - The only change to the online database is in the area of data submissions.

The on-line version of the ISFDB1 continues to have two distinct modes of operation:

- **Search and Display** - Users have the ability to search by title, author, year, or series. Awards for a specific year can be observed. An alphabetized author directory exists which the user can browse through. Select magazines have checklist tables which the user can browse through.

- **Data Submissions** - In data submission mode, forms are available to add/correct authors, publications, and title information. Data submissions will be collected by the ISFDB editor and submitted to the ISFDB2 for integration. Trusted users who wish to immediately incorporate data into the ISFDB2 will be required to use the user interface available at that site. Data submissions generated at the ISFDB1 will not be immediately integrated into the ISFDB2 for two reasons:

  - There is a general lack of knowledge and training amongst the general ISFDB population. This necessitates the use of an editor to insure that data conforms to ISFDB2 standards.

  - The data presented at the ISFDB1 may lack specific details present in the ISFDB2. For instance, page numbers where titles appear in a work are not displayed at the ISFDB1, but will be valid input at the ISFDB2.

5.4.2 ISFDB2

The ISFDB2 has four major modes of operation. These are:

- **The ISFDB2 User Interface**. Users with ISFDB2 accounts may log into the ISFDB2 system, and directly submit data into the database. While the specifics of the account administration is beyond the scope of this document, it is intended that these users are somehow trained and trusted in entering correctly formated information. Accountabily and change logs will enable erroneous data to be removed.

- **Upload of Change Logs**. The ISFDB2 will have a method of uploading change logs. These may be uploaded by the ISFDB1 editor on behalf of the users there, or they can be uploaded by users who have been making changes using the private off-line tools, or they can be submitted directly at the ISFDB2 web site. These change requests are handled by the change manager, which insures that the submitted data conforms to the formating requirements of the data repository, integrates the changes, and deposits a copy of the accepted changes in an official change log.

- **Download of Change Logs**. Synchronization of the off-line private tools is a two-way street. Not only do user changes need to be uploaded, by official changes need to be downloaded and applied to those off-line versions. In order to reduce the amount of necessary bandwidth need to stay synchronized, complete versions of the database will be available as a major release, while the change logs will represent point releases. In order to get completely up to date, a user will have to obtain the latest major release, and apply all of the point release change logs.

- **Download of Database**. As implied above, users will have the need to download the current released version of the database.

5.4.3 Offline Database

The offline version of the database should be of sufficient utility that users need not connect to the ISFDB2 unless they need to synchronize data. Otherwise, the offline version should contain the necessary search, display, and modification tools needed to make it useful.

The tools and data can be made available on CD-ROM for those who do not have sufficient bandwidth to download them from the ISFDB2 web site. The CD-ROM would be based off of the most recent major release; the tools would extract the latest point updates from the website to make the release current.

The tools should be written in a portable language (say Python) such that they can run either in the Windows, UNIX, or Mac environment.

5.4.4 Data Converters

Any number of data converters are possible. These converters would transform the XML data into some format usable by another tool such as Microsoft Access, or the ISFDB1. Only the ISFDB1 data converter is required prior to rollout.

5.5 User Classes and Other Involved Personnel

The user classes of the ISFDB remained unchanged from those specified in section 3.5.

5.5.1 Organizational Structure

Like most open source projects, the envisioned organizational structure is non-heirarchical. Someone, of course, must hold the keys in order to keep the vagrants out, but otherwise the running of the ISFDB should be a democratic process. Nonetheless, some partitioning of access is required:

- Only select individuals will be permitted to update the web site (although a Wiki-style website could be interesting).

- Users who wish to directly modify the ISFDB2 will be required to authenticate at the web site. Anonymous users will not be allowed to alter any data, although they may submit requests via the ISFDB1.

5.5.2 Profiles of User Classes

ISFDB maintainers - ISFDB maintainers should be able to perform all of their tasks with the offline tools, except during synchronization.

Authors/Bibliographers - Either the ISFDB1 or the ISDB2 can be utilized by this user class. Bibliographers may prefer using the ISFDB2, as more detailed data will be stored at that location. Most authors will probably desire to continue using the ISFDB1 due to its ease of use.

Fen/Readers - Casual readers and those who wish to quickly look up a title should continue using the ISFDB1, as the website was designed specifically for that purpose.

Bibliographic Linkers - User who wish to link to a persistant ISFDB bibliography should continue to utilize the ISFDB1 for this purpose.

Booksellers - Booksellers and collectors will probably desire to work exclusively from the offline version of the database, as it was designed specifically with them in mind.

5.5.3 Interactions Among User Classes

As no forums for communiation are present at the ISFDB, the only conduit for interaction among user classes is in the data deposited and integrated. Interaction may occur in other venues (such as USENET newsgroups), but these interactions are not controlled or monitored by ISFDB personnel.

5.5.4 Other Involved Personnel

None.

5.6 Support Environment

Support issues are left solely to the hosting ISP.

Major releases of the ISFDB2 will be found at the website, archived on CD-ROM, and may be available to the public on media as well. Point releases of the ISFDB2 will be found at the website.

6. Operational Scenarios

6.1 ISFDB1 Users

Operational scenarios for ISFDB1 users will be no different than they are on the existing system. For additional details, refer to the ISFDB FAQ at **http://www.sfsite.com/isfdb/faq.html**.

6.2 ISFDB2 Users

There are two types of users for the ISFDB2:

- **Synchronization** - These users are primarily offline database users, and only connect to the ISFDB2 website to perform synchronization. In this case, the synchronization function is built into the offline database tools, and the synchronization itself should be invisible to the user.

- **User Interface** - Some users may wish to input data directly into the ISFDB2 via its user interface. This user interface will be fairly limited, allowing users to:

    - Search for a specific title, author, series, or canonical record.

    - Allow the user to submit modifications to a title, author, series, or canonical record.

    - Allow the user to submit new title, author, series, or canonical records.

6.3 Offline Users

This is an entirely new operational scenario not found in the existing system. There are four major modes of operation:

- **Search** - The offline database will have the ability to perform some set of searching. Will it is desireable to perform complex searches on any database field, there may be practical limitations. Since the database consists of a large set of XML files, it could take several minutes to open each of the several thousand files while searching. A real searching facility may require the constructing of indices, but the operation of such indices would be compilcated by user modifications to data. At any rate, some minimal amount of searching is required (otherwise the database will have no utility), and a maximal amount of searching is desired.

- **Update** - The offline database will have the ability to make modifications to the database. These may be forms that are filled out (not unlike those at the present ISFDB), or may be presented in some other convenient method such as a spreadsheet.

- **Synchronization (Download)** - In order to sync up to the latest version of the data, users will need to synchronize with the ISFDB2 website. This process should be painless to the user, consisting of no more than activating their connection to the Internet, and pressing the "synchronize" button, or whatever facilty will exist for this purpose. The software will connect to the ISFDB2 website, and checks that offline version is using the same major release as the website. If not, the software will download the current major release. Once the major release is synchronized, it downloads all availble point releases and merges them.

- **Synchronization (Upload)** - After the user makes alterations to the offline version of the database, it will be desirable for the user to upload those changes to the website, so that everyone else can share them. During the process of making a change to a record, the editing tools will create an XML request which describes the record which was modified, the fields which were affected, and the specific changes that were made. At synchronization time, these requests are collected together and uploaded to the website. In order to minimize connect time (as well as simplifying the merging algorithms), only the requests will be uploaded; the rst of the database will not be transmitted.

6.4 ISFDB1 Editor

The operational scenarios for the ISFDB1 are greatly simplified under the proposed system. When an update to the website is desired, the ISFDB1 editor will perform the following steps:

- Synchronize his/her offline database to the latest version available at the ISFDB2 website. This step is identical to that of Offline Users in section 6.2 above.

- Using a data converter, convert the synchronized offline database into ISFDB1 format.

- Edit any static webpages that merit update, such as the What's New section, or the addition of new magazines to the checklist pages.

- Perform the usualy database build, and upload the result to the ISFDB1 website.

6.5 ISFDB2 Editor

The duties of the ISFDB2 editor will depend greatly on the final architecture of the site. Since specific scenarios are difficult to articulate at this time, and impact is limited to a single individual, we'll not worry about this use case at this time.

7. Summary of Impacts

7.1 Operational Impacts

For users of the ISFDB1, the operational impacts will be insignificant. Editors of the ISFDB will be able to share work more easily, posess the ability to edit offline, and find the merging of data more practical.

7.2 Organizational Impacts

The primary organizational impact is that an organization will exist where in the past it did not.

7.3 Impacts During Development

7.3.1 ISFDB1 Maintenance

During the period of time that the ISFDB2 is boot-strapped, there will be a period of time whereby ISFDB1 staff will have to devote some time to the creation of the ISFDB2. This will result in a slowdown in the integration of user-submitted data.

8. Analysis of the Proposed System

8.1 Summary of Improvements

The ISFDB2 will establish a true open data project for those interested in the generation and maintenance of speculative fiction bibliographies. It will provide a location and methodology to make this bibliographic data readily to everyone, and serve as a focal point for the development of tools for maintaining and searching the data. It will give users the ability to access the data without accessing the web page via the Internet, such that data entry can occur in libraries and private homes. Additionally, the new structure of the data should help to make the data itself more accurate.

8.2 Disadvantages and Limitations

The more open nature of the data can lead to its corruption by vandals and hackers. Limited restrictions are required to reduce the possiblity, and rollback/restoration procedures must be put in place to recover from such actions.

8.3 Alternatives and Trade-Offs Considered

A closed system utilizing a professional database would be easier to design and implement, but in the end would be less useful for the community at large.

9. Notes

This page intentionally left partially blank.

10. Appendices

This section contains example DTD and XML files. They are presented only to give a flavor for the intention described in this document. They are in no way to be considered the proposed or adopted formats which will be adopted by the ISFDB2.

The Binding Types, Content Types, and Work Types are compatible with those utilized by Contento and Miller in reference [7]. Regardless of what might be found in the examples here, final currency codes should be compatible with ISO 4217 (see reference [8]). Language codes should be compatible with ISO 639 2-letter codes (see reference [9]).

10.1 Author Files

10.1.1 Example Author DTD

```
<?xml version="1.0" encoding="UTF-8" ?>

<!-- #################################################
     author.dtd, version 0.1
     ################################################# -->


<!ENTITY % String        "(#PCDATA)" >

<!-- #################################################
     NON-TERMINAL ELEMENTS
     ################################################# -->

<!ELEMENT Author (
        Canonical,
        Actual,
        Birthdate?,
        Deathdate?,
        Country?,
        Pseudonyms?,
        Biography?,
        Emails?,
        Webpages?)>

<!ELEMENT  Pseudonyms  (Name)+ >
<!ELEMENT  Emails      (Email)+ >
<!ELEMENT  Webpages    (Webpage)+ >
<!ELEMENT  Webpage     (Url, Note?) >

<!-- #################################################
     TERMINAL ELEMENTS
     ################################################# -->

<!ELEMENT  Actual     %String; >
<!ELEMENT  Biography  %String; >
<!ELEMENT  Birthdate  %String; >
<!ELEMENT  Canonical  %String; >
<!ELEMENT  Country    %String; >
<!ELEMENT  Deathdate  %String; >
<!ELEMENT  Email      %String; >
<!ELEMENT  Name       %String; >
<!ELEMENT  Note       %String; >
<!ELEMENT  Url        %String; >
```

10.1.2 Example Author XML File

```
<?xml version='1.0' encoding='ascii'?>

<!DOCTYPE Publication SYSTEM "../../dtds/author.dtd">

<Author>
   <Canonical>Stephen Baxter</Canonical>
   <Actual>Baxter, Stephen M.</Actual>
   <Birthdate>1957</Birthdate>
   <Country>UK</Country>
   <Webpages>
      <Webpage>http://www.sam.math.ethz.ch/~pkeller/Baxter-Page.html</Webpage>
   </Webpages>
</Author>
```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

10.2 Publication Files

10.2.1 Example Publication DTD

```
<?xml version="1.0" encoding="UTF-8" ?>

<!-- #####################################################
     publication.dtd, version 0.1
     #####################################################  -->

<!-- #####################################################
     BINDING TYPES
     #####################################################
     unknown = Binding type not known
        Book Bindings:
          pb = paperback
          ph = pamphlet/chapbook
          hc = hardcover
          qp = quality paperback
          tb = tabloid
          tp = trade paperback
        Magazine Bindings:
          dg     = digest size magazine
          large  = 8.5 x 11"
          octavo = 5.5 x 8.5 saddle-stapled
          quarto = 8.4 x 11" saddle-stapled
     #####################################################  -->
<!ENTITY % binding_type     "pb|ph|hc|qp|tb|tp|dg|digest|
                             large|octavo|quarto|unknown">


<!-- #####################################################
     CONTENT TYPES
     #####################################################
          gn = graphic novel
          n  = novel
          om = omnibus (a collection of publications)
          an = anthology
          ga = graphic anthology
          co = collection of stories by one author
          gc = graphic collection
          mg = magazine
          gm = gamebook
          nf = non-fiction
          ng = non-genre
     #####################################################  -->
<!ENTITY % content_type     "gn|n|om|an|ga|co|gc|mg|gm|nf|ng">


<!-- #####################################################
     CURRENCY TYPES
     #####################################################
          AU  = Australian Dollars
          BR  = British Pounds
          CA  = Canadian Dollars
          DM  = Deutsch Marks
          IT  = Italian Lire
          US  = US Dollars
     #####################################################  -->
<!ENTITY % currency_type    "US|AU|UK|unknown">


<!-- #####################################################
     WORK TYPES
     #####################################################
```

```
                               ad = adaptation
                               ai = author's introduction to story
                               ar = article
                               aw = afterword
                               bg = biography
                               bi = bibliographic material
                               br = book review
                               cl = column
                               cr = criticism
                               cs = comic strip
                               ct = cartoon
                               ed = editorial
                               ex = extract
                               es = essay of unknown specific type
                               fa = facetious article
                               fw = foreword
                               gr = game review
                               gs = graphic story
                               hu = humor
                               il = illustration
                               in = introduction
                               ins= insert
                               is = introduction to story
                               iv = interview
                               ix = index
                               lk = linking material
                               lt = letter
                               mr = movie review
                               ms = miscellaneous
                               n  = novel
                               na = novella
                               nv = novelette
                               pi = pictorial
                               pl = play
                               pm = poem
                               pp = prose poem
                               pr = preface
                               pt = photography
                               pz = puzzle
                               qz = quiz
                               rr = round-robin
                               sa = story adaptation of a play/screenplay
                               sf = special feature
                               sg = song
                               si = section introduction
                               sl = serial segment
                               sp = speech
                               sr = story review
                               ss = short story
                               tc = title change
                               tr = translation
                               uw = unfinished work
                               vi = vignette
                    ################################################## -->
              <!ENTITY % work_type    "ad|ai|ar|aw|bg|bi|br|cl|cr|cs|ct|ed|es|
                                       ex|fa|fw|gr|gs|hu|il|in|ins|is|iv|ix|
                                       lk|lt|mr|ms|na|nv|pi|pl|pm|pp|pr|
                                       pt|pz|qz|rr|sa|sf|sg|si|sl|sp|sr|ss|
                                       tc|tr|uw|vi|n">

              <!ENTITY % String         "(#PCDATA)" >

              <!-- ##################################################
                   NON-TERMINAL ELEMENTS
                   ################################################## -->
```

```
 1              <!ELEMENT        Publication              (Title,
 2                                                        (Author|Authors),
 3                                                        Date,
 4                                                        Identifiers?,
 5                                                        Price?,
 6                                                        Pages?,
 7                                                        Publisher?,
 8                                                        CoverTitle?,
 9                                                        (CoverArtist|CoverArtists)?,
10                                                        BackCoverTitle?,
11                                                        (BackCoverArtist|BackCoverArtists)?,
12                                                        Note?,
13                                                        AnnualTitle?,
14                                                        History,
15                                                        Content?
16                                                        )>
17
18              <!ELEMENT        Authors                  (Author)+
19                                                        >
20              <!ELEMENT        Identifiers              (UniqueID,
21                                                        (ISBN|ISSN|LCC|Dewey|Catalog)*
22                                                        )>
23              <!ELEMENT        CoverArtists             (CoverArtist)+
24                                                        >
25              <!ELEMENT        BackCoverArtists         (BackCoverArtist)+
26                                                        >
27              <!ELEMENT        History                  (Created,
28                                                        (Modified)*,
29                                                        (Verified)*
30                                                        )>
31              <!ELEMENT        Content                  (Empty | (Work)+)
32                                                        >
33
34
35              <!ELEMENT        Work                     (Title,
36                                                        (Author|Authors)?,
37                                                        Origin?,
38                                                        Translation?,
39                                                        (Reviewer|Reviewers)?,
40                                                        (Interviewer|Interviewers)?)
41                                                        >
42              <!ELEMENT        Translation              (Language?,
43                                                        Date?,
44                                                        (Author|Authors)
45                                                        )>
46              <!ELEMENT        Reviewers                (Reviewer)+
47                                                        >
48              <!ELEMENT        Interviewers             (Interviewer)+
49                                                        >
50
51              <!-- ###################################################
52                   TERMINAL ELEMENTS
53                   ###################################################  -->
54              <!ELEMENT        AnnualTitle              %String; >
55              <!ELEMENT        Author                   %String; >
56              <!ELEMENT        BackCoverArtist          %String; >
57              <!ELEMENT        BackCoverTitle           %String; >
58              <!ELEMENT        Catalog                  %String; >
59              <!ELEMENT        CoverArtist              %String; >
60              <!ELEMENT        CoverTitle               %String; >
61              <!ELEMENT        Created                  %String; >
62              <!ELEMENT        Date                     %String; >
63              <!ELEMENT        Dewey                    %String; >
64              <!ELEMENT        Empty                    %String; >
65              <!ELEMENT        Interviewer              %String; >
```

```
1    <!ELEMENT        ISBN              %String; >
2    <!ELEMENT        ISSN              %String; >
3    <!ELEMENT        Origin            %String; >
4    <!ELEMENT        Modified          %String; >
5    <!ELEMENT        Language          %String; >
6    <!ELEMENT        LCC               %String; >
7    <!ELEMENT        Note              %String; >
8    <!ELEMENT        Pages             %String; >
9    <!ELEMENT        Price             %String; >
10   <!ELEMENT        Publisher         %String; >
11   <!ELEMENT        Reviewer          %String; >
12   <!ELEMENT        Title             %String; >
13   <!ELEMENT        UniqueID          %String; >
14   <!ELEMENT        Verified          %String; >


     <!-- ####################################################
            ELEMENT ATTRIBUTES
            #################################################### -->
     <!ATTLIST  Content     type        (%content_type;)      #IMPLIED>
     <!ATTLIST  Price       currency    (%currency_type;)     #IMPLIED>
     <!ATTLIST  Publication binding     (%binding_type;)      #IMPLIED>
     <!ATTLIST  Work        page        CDATA                 #IMPLIED
                            type        (%work_type;)         #IMPLIED
                            copyright   CDATA                 #IMPLIED>
```

10.2.2 Example Publication XML File

```
<?xml version='1.0' encoding='ascii'?>

<!DOCTYPE Publication SYSTEM "../../../dtds/publication.dtd">

<Publication binding="unknown">
   <Title>Weird Tales, March 1923</Title>
   <Author>Edwin Baird</Author>
   <Date>1923</Date>
   <Identifiers>
      <UniqueID>WEIRDTMAR1923</UniqueID>
   </Identifiers>
   <Price currency="US">0.25</Price>
   <Pages>192</Pages>
   <CoverArtist>R. R. Epperly</CoverArtist>
   <Note>Vol. 1 #1</Note>
   <AnnualTitle>Weird Tales - 1923</AnnualTitle>
   <History>
      <Created>Generated from ISFDB 07/12/2001.</Created>
   </History>
   <Content type="mg">
      <Work type="nv">
         <Title>Ooze</Title>
         <Author>Anthony M. Rud</Author>
      </Work>
      <Work type="ss">
         <Title>The Chain</Title>
         <Author>Hamilton Craigie</Author>
      </Work>
      <Work type="ss">
         <Title>The Ape-Man</Title>
         <Author>James B. M. Clark, Jr.</Author>
      </Work>
      <Work type="ss">
```

```
 1              <Title>The Skull</Title>
 2              <Author>Harold Ward</Author>
 3            </Work>
 4            <Work type="ss">
 5              <Title>The Gallows</Title>
 6              <Author>I. W. D. Peters</Author>
 7            </Work>
 8            <Work type="ss">
 9              <Title>The House of Death</Title>
10              <Author>F. Georgia Stroup</Author>
11            </Work>
12            <Work type="ss">
13              <Title>The Return of Paul Slavsky</Title>
14              <Author>George Warburton Lewis</Author>
15            </Work>
16            <Work type="ss">
17              <Title>The Extraordinary Experiment of Dr. Calgroni</Title>
18              <Authors>
19                 <Author>Joseph Faus</Author>
20                 <Author>James Bennett Wooding</Author>
21              </Authors>
22            </Work>
23            <Work type="ss">
24              <Title>The Scarlet Night</Title>
25              <Author>William Sanford</Author>
26            </Work>
27            <Work type="ss">
28              <Title>The Young Man Who Wanted to Die</Title>
29              <Author>Anonymous</Author>
30            </Work>
31            <Work type="ss">
32              <Title>Nimba, the Cave Girl</Title>
33              <Author>R. T. M. Scott</Author>
34            </Work>
35            <Work type="ss">
36              <Title>The Weaving Shadows</Title>
37              <Author>W. H. Holmes</Author>
38            </Work>
39            <Work type="ss">
40              <Title>The Sequel</Title>
41              <Author>Walter Scott Story</Author>
42            </Work>
43            <Work type="nv">
44              <Title>The Accusing Voice</Title>
45              <Author>Meredith Davis</Author>
46            </Work>
47            <Work type="ss">
48              <Title>The Basket</Title>
49              <Author>Herbert J. Mangham</Author>
50            </Work>
51            <Work type="ss">
52              <Title>The Unknown Beast</Title>
53              <Author>Howard Ellis Davis</Author>
54            </Work>
55            <Work type="ss">
56              <Title>The Closing Hand</Title>
57              <Author>Farnsworth Wright</Author>
58            </Work>
59            <Work type="ss">
60              <Title>The Place of Madness</Title>
61              <Author>Merlin Moore Taylor</Author>
62            </Work>
63            <Work type="ss">
64              <Title>Fear</Title>
65              <Author>David R. Solomon</Author>
```

```
                              </Work>
                              <Work type="ss">
                                 <Title>The Ghoul and the Corpse</Title>
                                 <Author>G. A. Wells</Author>
                              </Work>
                              <Work type="ss">
                                 <Title>The Ghost Guard</Title>
                                 <Author>Bryan Irvine</Author>
                              </Work>
                              <Work type="ss">
                                 <Title>Hark! The Rattle!</Title>
                                 <Author>Joel Townsley Rogers</Author>
                              </Work>
                              <Work type="ss">
                                 <Title>The Grave</Title>
                                 <Author>Orville R. Emerson</Author>
                              </Work>
                              <Work type="ss">
                                 <Title>The Mystery of Black Jean</Title>
                                 <Author>Julian Kilman</Author>
                              </Work>
                              <Work type="ss">
                                 <Title>The Dead Man's Tale</Title>
                                 <Author>Willard E. Hawkins</Author>
                              </Work>
                              <Work type="sl">
                                 <Title>The Thing of a Thousand Shapes (Part 1 of 2)</Title>
                                 <Author>Otis Adelbert Kline</Author>
                              </Work>
                           </Content>
                        </Publication>
```

10.3 Canonical Files

10.3.1 Example Canonical DTD

```xml
<?xml version="1.0" encoding="UTF-8" ?>

<!-- ####################################################
     canonical.dtd, version 0.1
     #################################################### -->

<!-- ####################################################
     CONTENT TYPES
     ####################################################
         n  = novel
         om = omnibus (a collection of publications)
         an = anthology
         co = collection of stories by one author
         nf = non-fiction
         ng = non-genre
     #################################################### -->

<!ENTITY % content_type    "n|om|an|co|nf|na|nv|ss|es|ng|sl|pm|ee">

<!ENTITY % String          "(#PCDATA)" >

<!-- ####################################################
     NON-TERMINAL ELEMENTS
     #################################################### -->

<!ELEMENT Work (
        Title,
        (Author|Authors),
        Series?,
        SeriesNum?,
        Synopsis?,
        Notes?,
        Awards?,
        Variant* )>

<!ELEMENT  Authors       (Author)+ >
<!ELEMENT  Variant       ((Author|Authors) | Title )+ >
<!ELEMENT  Awards        (Award)+ >

<!-- ####################################################
     NON-TERMINAL ELEMENTS
     #################################################### -->

<!ELEMENT Work (
        Title,
        (Author|Authors),
        Series?,
        SeriesNum?,
        Synopsis?,
        Notes?,
        Awards?,
        Variant* )>

<!ELEMENT  Authors       (Author)+ >
<!ELEMENT  Variant       ((Author|Authors) | Title )+ >
<!ELEMENT  Awards        (Award)+ >

<!-- ####################################################
     TERMINAL ELEMENTS
     #################################################### -->
```

```
<!ELEMENT   Title        %String; >
<!ELEMENT   Author       %String; >
<!ELEMENT   Synopsis     %String; >
<!ELEMENT   Name         %String; >
<!ELEMENT   Notes        %String; >
<!ELEMENT   Number       %String; >
<!ELEMENT   Series       %String; >
<!ELEMENT   SeriesNum    %String; >
<!ELEMENT   Award        %String; >


<!-- ##################################################
     ELEMENT ATTRIBUTES
     ##################################################  -->

<!ATTLIST   Work         type        (%content_type;)        #IMPLIED
                         copyright   CDATA                   #IMPLIED>
```

10.3.2 Example Canonical XML File

```
<?xml version='1.0' encoding='ascii'?>

<!DOCTYPE Work SYSTEM "../../dtds/canonical.dtd">

<Work type="n" copyright="1990">
   <Title>Use of Weapons</Title>
   <Author>Iain M. Banks</Author>
   <Series>Culture</Series>
   <SeriesNum>4</SeriesNum>
   <Awards>
      <Award>787</Award>
   </Awards>
</Work>
```

11. Glossary

ASCII	American Standard Code for Information Interchange
fen	Plural of fan
ISFDB	Internet Speculative Fiction Database
HTML	HyperText Markup Language
ISFDB	Internet Speculative Fiction Database
ISFDB1	The original ISFDB, located at *http://www.sfsite.com/isfdb*
ISFDB2	Data entry adjunct to the ISFDB, located at *http://www.isfdb.org*
ISP	Internet Service Provider
metadata	Information about a publication (such as ISBN, page count, etc).
pipe delimter	When using plain-text files to store data, the fields are often separated by a special character. Many systems use the "|" symbol as a field delimeter, as it is otherwise seldom used. This reduces the chance of misinterpretation of data as a field delimeter.