# DAISY Consortium logoDAISY Consortium

Registered as an Association under the civil code of Switzerland

Jesper Klein, President - Chairman of the Board

President@daisy.org

[www.daisy.org](http://www.daisy.org)

**Proposal for EpubCheck Development Update and Maintenance**

# Company overview

## Principal contact for this proposal:

Avneesh Singh

Chief Operating Officer

[asingh@daisy.org](mailto:asingh@daisy.org)

## Our organization

The DAISY Consortium is a non-profit organization incorporated in Zurich, Switzerland, and operating worldwide. Our mission is to ensure that information and knowledge is useable and accessible for everybody, including persons with disabilities. See our website at:

<http://www.daisy.org>

Our organization is regulated by the Swiss laws of associations. We are governed by a board of directors appointed by our full members. We are financially secure and compliant with our capital reserves policy. Our accounts are audited annually by Betschon Treuhand AG, Zurich. Annual reports, including the financial summary, are available on our website. Our corporate bank accounts are with Credit Suisse, Zurich and First Interstate Bank, Missoula.

## Our History

The DAISY Consortium was formed in 1996 by libraries serving the blind and other persons with disabilities to find a replacement for the dying audio cassette format. In 1997, we embraced a standards-based approach to solve the problem and joined the W3C. We participated in development of the SMIL specification and in 1999 published the DAISY 2 Standard. That same year, we joined in the activities that lead to the formation of the Open eBook Forum (OEBF), which became the IDPF and is now combined with the W3C.

We soon recognized the need to have a method for validating content in the DAISY format; we developed the DAISY suite of validation tools. We were then able to test and promote interoperability between libraries and the commercial reading systems they used. We brought this concept with us to the OEBF and over time, this lead to what is today EpubCheck.

In 2011, the DAISY Consortium endorsed the EPUB 3 Standard and stopped all development of the older DAISY standard. We are encouraging all of our DAISY Members to adopt EPUB 3 and move towards integration with the mainstream. For these reasons, the DAISY Consortium feels strongly that a solid implementation of EpubCheck is critical for the digital publishing industry, which brings this work in scope for our Mission.

## Our People

We are a small company, but we punch like a heavyweight. There are fifteen of us spread throughout the world. We have staff in India, Canada, the United States, United Kingdom, Belgium, and France. We focus our work on what is strategically most effective. Our software developers also participate in the EPUB and WP Standards development, which we believe is important in responding to this RFP. We are all involved with the various aspects of EPUB validation and interoperability.

## Custom solutions from DAISY

The DAISY Consortium is a firm believer in open source software development. Over many years and multiple projects, we have developed approaches that enable maximum participation from the user and development communities. Our software projects are geared to help our more than 100 members produce and distribute conforming digital publications to their library patrons. We also run projects that are focused on helping to make mainstream publications accessible to persons with disabilities. Some of the current software projects we run or participate in are:

* We have actively been helping to maintain EpubCheck since the EPUB 2.0 days. We know that many developers have made contributions without a rigorous design, which has led to the current mix of coding approaches that need to be resolved in a future release.
* The DAISY Pipeline is software used for the automatic conversion from one format to another. We have promised our DAISY Consortium members that they will be able to move their digital books forward as the formats evolve. For example, we have conversions from DAISY to EPUB 3. This software uses Java, so we have many years of experience in Java development.
* Ace by DAISY is an accessibility checker for EPUB publications. This was launched in January of 2018 and has been adopted broadly in the publishing community. Ace is based on JavaScript and our developers now have experience in that software environment.
* Obi is an audio book authoring tool widely used throughout the world by DAISY Consortium members. This can produce DAISY titles, and EPUB 3 with media overlays.
* Tobi is a full text full audio with synchronization authoring tool used by DAISY organizations producing full text and audio titles. It too supports EPUB 3 using media overlays.

Note: All the software, interfaces, and websites we develop are fully accessible to persons with disabilities.

## Top clients and services provided

A long, long list of publishers and vendors to the digital publishing industry are using Ace by DAISY. Two specific examples are VitalSource in the US and Hachette Livre in France.

DAISY Consortium members all around the world are using Obi, Tobi and the DAISY Pipeline in their day-to-day work. Hundreds of thousands of titles are produced and converted by DAISY technology each year, in many languages. One specific example is the National Library for the Blind (NLB) in Norway, which not only utilizes tools developed by the DAISY Consortium, but also contributes actively to their development. A second example is Bookshare, which uses DAISY Pipeline for the processing of more than 600,000 titles.

## Awards and industry recognition.

In June 2018 the DAISY Consortium was awarded the CNIB Century of Change Award. DAISY was honored for its ongoing work in developing global solutions for accessible publishing and reading. “The work of the DAISY Consortium has directly contributed to the advancement of library and information access and has had a particularly profound impact on the lives of Canadians with sight loss.”

In 2016, having reviewed over 1000 submissions, the Google Foundation selected the DAISY Consortium to receive the Google Impact Challenge Award. The significant funding for our Inclusive Publishing initiative represented a huge vote of confidence in our ability to deliver a complex technical project on time and budget.

<https://www.prnewswire.com/news-releases/daisy-consortium-wins-grant-from-googleorg-300250002.html>

The DAISY Consortium was awarded the Andrew W. Mellon Foundation for Technical Collaboration (MATC)

<https://www.pr.com/press-release/63998>

The International Telecommunications Union (ITU), the leading United Nations agency for information and communication technology issues, awarded its World Telecommunications and Information Society Award 2008 to the DAISY Consortium.

<https://www.pr.com/press-release/85666>

## Work samples

Two ongoing open source projects are the most relevant for the purposes of this proposal:

1. DAISY Pipeline: Development site found at http://daisy.github.io/pipeline/
2. Ace by DAISY: Development site found at https://github.com/daisy/ace and the consumer-facing site at: https://inclusivepublishing.org/toolbox/accessibility-checker/

# Implementation

## Project management, tools and programming language

The existing tools will be continued:

* Java-based implementation, developers are free to use their code editor of choice
* Maven build tool
* Transifex service for translation management

GitHub issue tracker will be used extensively for managing & tracking the development:

* A revamped label organization will be proposed, with dedicated process documentation
* A milestone strategy will be defined, with dedicated process documentation, to prioritize issues and define target releases (at least one milestone for each target release)
* GitHub "projects" may be used for ad-hoc feature development

Formalize the contribution and implementation process, with dedicated documentation:

* Each new feature ("check") should be defined using a behavior-driven development (BDD) or test-driven development (TDD) approach: define the expected behavior with unit tests first, then implement. This ensures that all new feature development is fully tested and helps with the long-term quality and maintenance of the project.
* Each change to the code base is submitted first as pull request (PR), which should be reviewed before being merged in. Trusted maintainers may be able to exceptionally merge their own PRs.
* Branching patterns and branch name conventions are defined
* setup a contributor’s license agreement (CLA) signing process (for any future contributions, and if possible for the past contributors we can contact)

Facilitate direct and asynchronous communication among developers and possibly users too:

* establish weekly or bi-weekly developers calls (chat or audio)
* launch an open communication channel for developers (using internet-relay chat (IRC) or any Slack-like solution)
* if need be, launch an open communication channel for users (using internet-relay chat (IRC) or any Slack-like solution)

Help contributors and users feel at home in an inclusive environment:

* Add a code of conduct to the project, including a process to enforce its rules
* Review and possibly revise the "Welcoming" and "Contributing" documentation for newcomers
* If need be, provide issue and pull request templates

Update licensing information when this is agreed by W3C authorities

### Rationale for using Java programming language

* Java is the go-to language for XML processing, with the most correct libraries. For instance, the most complete and standard-compliant libraries for Schematron, XSLT, schema-based validation are Java based.
* The Nu HTML Checker is implemented in Java, so staying in Java will make the integration easier
* It saves the cost of porting to another language
* Java is still among the most used programming languages, and it is expected that Java programming skills will be reasonably available in longer term.

## Skill set offered by DAISY to effectively implement the solution

* excellent knowledge of EPUB specifications
* Java development skills
* familiarity with XML processing APIs (SAX, StAX, DOM)
* familiarity with schema languages (RelaxNG, Schematron, XML Schema.)
* knowledge of testing methodologies
* experience with the Maven build tool
* ability to read and understand technical specifications from IDPF and W3C
* proven approaches that encourage and facilitate community involvement in open source development
* fluent/professional in English (spoken and written)
* communication skills, especially with an international team across various time zones

In addition to deploying DAISY Consortium’s skilled developers, we also plan to invite contributions from experienced EpubCheck developers like Tobias Fischer on a paid basis.

## Feature development

### Phase 1: Maintenance release for important issues and updating EpubCheck to EPUB 3.2

#### Milestone 1.1: Release of EpubCheck 4.1

Envisioned Work items:

* wrap-up issue #789 and associated pull request #789
* review pull request #650, devise integration options and implement
* build, test, and release

Deliverables:

* EpubCheck v4.1.0 maintenance release

#### Milestone 1.2: Implementation of EPUB 3.2

Envisioned Work items:

* Identification of new checks required by EPUB 3.2
* create dedicated EpubCheck issues to track progress of each new check
* liaise with the EPUB 3 Community Group (CG)
* Define, together with the EPUB 3 CG and Publishing Business Group (BG), the processing logic for backwards compatibility (e.g. should EpubCheck be able to check a specific older version like 3.0.1 or just the latest backward-compatible 3.x).
* Creation of at least one passing and one failing test for each change identified above
* Propose implementations of each issue as a pull request
* includes patching the custom CSS parser to fix critical issues
* includes updating the XHTML schemas to introduce recent changes in HTML
* Review PRs
* Build, test, and release

Deliverables:

* At least one beta release
* At least one release candidate
* EpubCheck v4.2.0, updated to EPUB 3.2

### Phase 2: Clean-up, Documentation, Refactoring and Media Overlays support

#### Milestone 2.1: Test suite cleanup and refactoring

Envisioned Work items:

* clean up the test suite
* normalize and simplify the test samples (remove all the cruft, make them as atomic as possible)
* experiment with a BDD approach for tests
* in any case (BDD or not) refactor and reorganize the test suite to make it closer to the spec (and consequently, easier to look up and maintain)

Deliverables

* new test suite (Will be published as a maintenance release)

#### Milestone 2.2: New API, documentation and Media Overlays support

* Application programming interface (API) and messaging refactoring
* thorough review and simplification of the public API
* thorough review and consolidation of the messaging infrastructure (how to serialize messages in various flavors: text, JSON, XML; consolidate the INFO and USAGE messages, make sure they’re visible when needed)
* consolidation of the code base (notably removal/integration of the "ctc" package from the B&N contribution)
* Code documentation
* documentation of the public API
* documentation of the validation workflow
* documentation of the key classes and utilities (that are used when implementing new checks or features)
* Development of a public web site, for providing a user-friendly interface. It will mainly include:
  + Easy access to download the latest version
  + User documentation
  + pointers to contributing and helping guides
* Better checking of Media Overlays. It will include:
  + improve validation of durations and SMIL clock values
  + validation of reading order

Deliverables

* consolidated public API (will be published as a new release, v4.3.0 or v5.0.0 depending on the changes)
* public web site for EpubCheck
* EpubCheck with better Media Overlays support

### Phase 3: HTML checker prototype

The objective is to conduct a feasibility study for integrating HTML Checker in EpubCheck to evaluate the technical feasibility and estimate the development efforts for full-fledged integration.

Envisioned work plan:

* Integration of the Nu HTML Checker (the reference HTML checker, used by W3C’s checking service)
* investigate on the technical feasibility
* implement a prototype integration

Deliverables:

* A prototype EpubCheck "next generation" based off the Nu HTML Checker

Rationale for integrating HTML Checker:

EPUB 3.2 is using undated references to HTML, which means that EpubCheck has to be consistently updated for each new version of HTML. Integration of HTML Checker would significantly reduce the efforts for updating separate HTML Checker of EpubCheck and would provide more accurate results.

Furthermore, publication specifications in W3C are moving closer and closer to HTML and web technologies. WP specifications will be a major milestone on this trajectory. Therefore, the integration of HTML Checker would reduce the maintenance and development cost of EpubCheck in the long run.

### Phase 4: Integration of the Nu HTML Checker (To be defined)

This phase is included as a placeholder to indicate a possible future direction. It will be informed by the outcomes of Phase 3. This phase is not included in the timeline or costings.

Envisioned Work items:

* Integration of the Nu HTML Checker as prototyped in phase 3

Deliverables:

* A stable release of EpubCheck "next generation" (v5.0?)

The estimate of development efforts will be determined on the basis of Phase 3.

## Regular Maintenance

Maintenance will be a regular activity which will start with commencement of contract and will continue till the 6th month after completion of phase 2.

### Issue triage

Envisioned Work items:

* ensure a 48h response time for each newly posted issue (on business week days)
* reproduce issues, gather feedback and test samples
* sort issues with the proper GitHub label
* assign to a developer responsible for the fixing or implementation

### Bug fixes

Envisioned Work items:

* fix most critical bugs
* publish maintenance releases when needed

Deliverable(s):

* Maintenance releases v4.2.x

### Schema and vocabulary maintenance

EpubCheck uses schema for validation which was originally developed from schema of

EPUB 3 specifications. It also uses additional schema for advanced rules.

The schema of EPUB 3 specifications is not well maintained. Therefore, if W3C community Group wants to replace it with EpubCheck schema, a community group report can be created from EpubCheck schema, and maintained through the project timeline.

Envisioned Work items:

* track changes in the various EPUB vocabularies
* track changes in HTML and update EpubCheck schemas
* if the EPUB 3 CG needs, publish EpubCheck schemas as CG reports

### Scope management and provision of adjustments

* New feature requests will be considered during the duration of the project and will be prioritized according to the demand and availability of development resources.
* The envision work plan provides the tentative activities, which may need some adjustments based on practicalities.

# Contract and pricing

## Timeline and cost

The following timeline is proposed assuming that the project will start from mid-September 2018 and will continue till May 2020. If the project commencement is delayed, then the deliverables will be delayed accordingly.

### Phase 1: Maintenance release for important issues and updating EpubCheck to EPUB 3.2

[September 2018 to February 2019]

The month of September will be mainly focused on setting up infrastructure of the project. This will also provide a buffer of couple of weeks if the project is not able to officially commence from mid-September.

#### Milestone 1.1: Release of EpubCheck 4.1

Timeline:

Maintenance release: November 2018

#### Milestone 1.2: Implementation of EPUB 3.2

Timeline:

Start preview releases or alpha: Mid December 2018

Start testing stage, with release of beta: January 2019

Stable EpubCheck updated for EPUB 3.2: February 2019

Cost of development and maintenance for phase 1: US$ 25,800

### Phase 2: Clean-up, Documentation, Refactoring and Media Overlays support

[February to November 2019]

#### Milestone 2.1: Test suite cleanup and refactoring

Timeline:

New test suite: June 2019

#### Milestone 2.2: New API, documentation and Media Overlays support

Consolidated public API, new website and Media Overlays support: November 2019

Cost for development and maintenance for phase 2: **US$ 79,200**

### Phase 3: HTML checker prototype

[December 2019 to February 2020]

Timeline:

Prototype build on HTML Checker: February 2020.

Cost for development for phase 3: **US$ 13,100**

### Phase 4: Integration of the Nu HTML Checker (To be defined)

When the findings of the feasibility study in Phase 3 are completed, we will provide estimated timelines and costings for full-fledged integration of the HTML Checker.

### Follow-up Maintenance

[December 2019 to May 2020]

Timeline:

At least one maintenance release in May 2020.

Additional maintenance releases can be provided between December 2019 and May 2020 as per the requirements.

Cost for maintenance: **US$ 18,400**

Total development and maintenance cost: **US$ 136,500**

## Proposed contract terms

**Communications:** W3C consents and authorizes DAISY Consortium to communicate its role in developing and maintaining the software. Specifically, DAISY Consortium may state that “EpubCheck is maintained by DAISY”.

**Contract period:** The contract will commence on September 15, 2018 and end May 31, 2020.

**Payment:** Each quarter will be invoiced in advance. Invoices will be sent by e-mail. Invoices are due within 30 days of the invoice date.

# Customer references

**1. Rick Johnson**VP of Product StrategyVitalSource Technologies Inc  
227 Fayetteville St #400  
Raleigh, NC 27601, USA   
Telephone: +1 919-755-8100  
email: Rick.Johnson@ingramcontent.com

**2. Luc Audrain**  
Head of Digitalization  
Hachette Livre  
11 Rue Paul Bert, 75011 Paris, France  
Telephone: +33 (0) 6 48 38 21 41  
email: LAUDRAIN@hachette-livre.fr

**3. Arne Kyrkjebø**Deputy Director Development Department  
The Norwegian Library of Talking Books and Braille (NLB)  
Observatoriegaten 1B  
Box 2764  
Solli 0204   
Oslo, Norway  
email: arne.kyrkjebo@nlb.no