The NIH Manuscript Submission System

Abigail Acland[⊠][1]

Created: March 15, 2013.

Summary

The NIH Manuscript Submission system (NIHMS) handles the submission of full-text manuscript material to PubMed Central (PMC) in support of the NIH Public Access and other related funding agency policies for material published in non-PMC participating journals. The NIHMS system allows users, such as authors, principal investigators (PIs), and publishers to supply material for conversion to XML documents in a format that can be ingested by PMC. Documents go through multiple stages in the conversion process; initial submission, grant reporting, initial conversion approval, staff QA, conversion, QA of converted material, Web version approval, and citation matching.

History

NIHMS was created in 2005 as part of the NIH Public Access Policy. The policy calls for deposit of NIH funded research in PubMed Central. For authors whose articles were published in journals that are not submitting content to PMC directly, a method for submission was needed. NIHMS was created to allow users to deposit manuscript source files for conversion to PMC articles.

Author Affiliation: 1 NCBI; Email: aclanda@ncbi.nlm.nih.gov.

⊠ Corresponding author.

Overview of Submission Methods

	Method A Journal deposits final published articles in PubMed Central without author involvement	**Method B** Author asks publisher to deposit specific final published article in PMC	**Method C** Author deposits final peer-reviewed manuscript in PMC via the NIHMS	**Method D** Author completes submission of final peer-reviewed manuscript deposited by publisher in the NIHMS
Version of Paper Submitted	Final Published Article	Final Published Article	Final Peer-Reviewed Manuscript	Final Peer-Reviewed Manuscript
Task 1: Who starts the deposit process?	Publisher	Publisher	Author or designee, via NIHMS	Publisher
Task 2: Who approves paper for processing?	Publisher	Publisher	Author, via NIHMS	Author, via NIHMS
Task 3: Who approves paper for Pub Med Central display?	Publisher	Publisher	Author, via NIHMS	Author, via NIHMS
Participating journal/publisher	Method A Journals	Make arrangements with these publishers	Check publishing agreement	Make arrangements with these publishers
Who is Responsible?	NIH Awardee	NIH Awardee	NIH Awardee	NIH Awardee
To cite papers, from acceptance for publication to 3 months post publication	PMCID or "PMC Journal- In Process"	PMCID or "PMC Journal- In Process"	PMCID or NIHMSID	PMCID or NIHMSID
To cite papers, 3 months post publication and beyond	PMCID	PMCID	PMCID	PMCID

The policy was initially voluntary and took affect in May 2005. Compliance remained low in the months following. The policy then became mandatory in April 2008.

Ingest
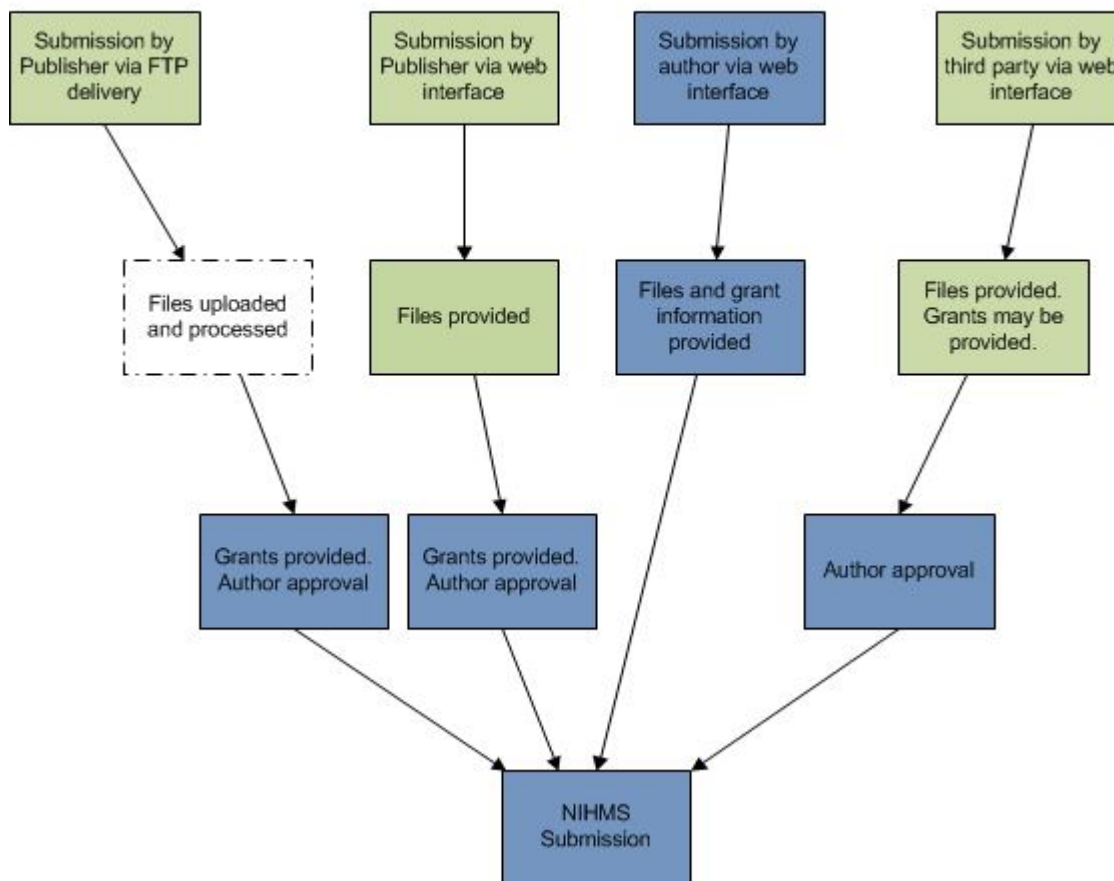
NIHMS receives source files from a variety of submitters. The majority of submissions are entered as single manuscripts through a Web interface. Submissions may be started by anyone with an account in the system, but they must be approved by an author on the paper, and appropriate funding must be reported for the record.

A high percentage of manuscripts in NIHMS are started by publishers as a service to their authors.

Ingest QA

Once the material has been submitted to NIHMS, it must undergo a QA review by staff to ensure that the material is complete, suitable for conversion, within scope of the NIH Public Access Policy, and does not represent a duplicate submission. Staff also checks funding to ensure reported funding is applicable and that the approving party on the submission is an author on the paper.

Submissions that pass the staff QA evaluation will be sent on in processing for conversion. Items that do not pass QA may be blocked from further processing, merged with an existing record, or returned to the submitter or author for correction.

Additional automated checks are performed on the submission at this stage to ensure that the material is matched to a journal recognized by the NLM and does not fall within any PMC-participating journal's period of

PMC participation. PIs holding associated funding who were not involved in the submission process are notified of the deposit of material associated with their grants at this time.

Conversion

Once these requirements have been met, the manuscript is sent to document conversion vendors for conversion to XML.

The conversion vendors return to NIHMS the full-text XML file for the manuscript, figure files in standard formats for thumbnail, Web, and PDF presentation, and any extracted supplemental files that may have originally been provided as embedded in the original source files.

XML in NIHMS

Regardless of source document format, all submitted manuscripts are converted to full-text XML using the NLM JATS (Journal Article Tagging Suite) Journal Publishing model.

The XML document is considered the converted document. The Web and PDF displays rendered from the XML are display versions of the document, but alterations and corrections are performed on the base XML or rendering software, rather than directly on the Web or PDF version.
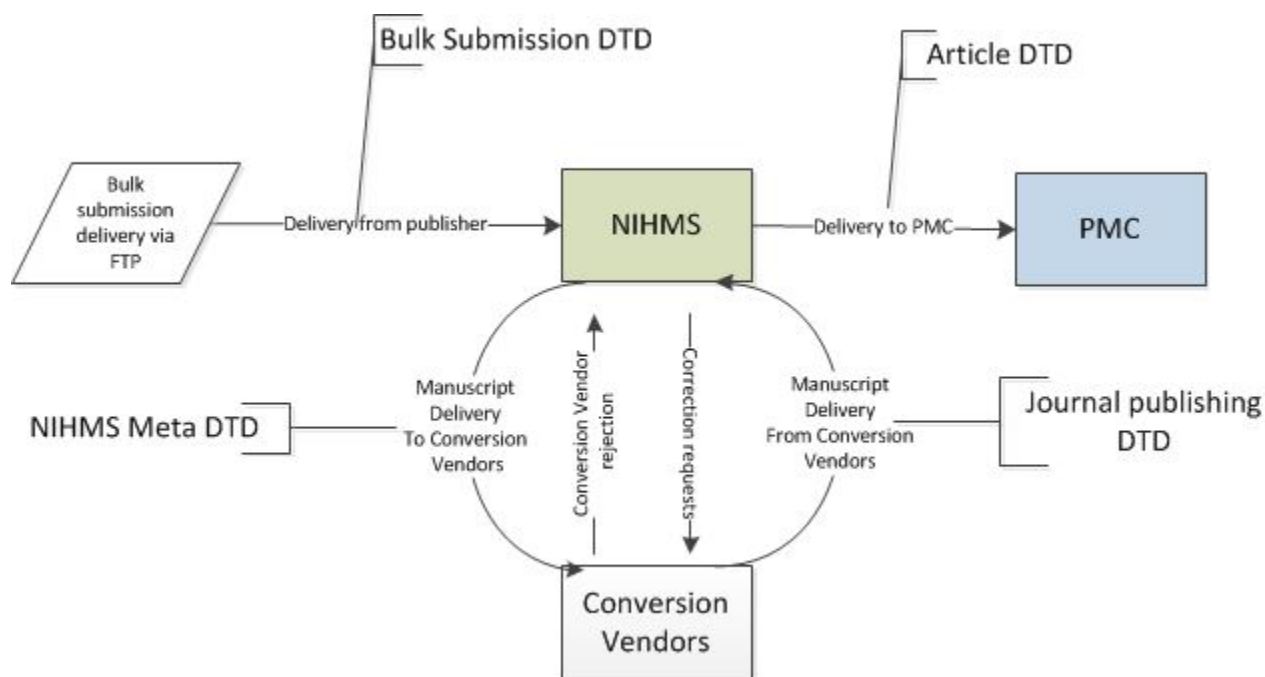
Complex mathematical content that cannot be captured in plain JATS markup and display formulas are captured using MathML.

Source File	$p_\lambda(t) = \frac{1}{2}[\lambda^2 - (\lambda - t)_+^2]$

MathML	`<mml:math id="M19" display='block'>`
	`<mml:msub>`
	`<mml:mi>p</mml:mi>`
	`<mml:mi>λ</mml:mi>`
	`</mml:msub>`
	`<mml:mo stretchy='false'>(</mml:mo>`
	`<mml:mi>t</mml:mi>`
	`<mml:mo stretchy='false'>)</mml:mo>`
	`<mml:mo>=</mml:mo>`
	`<mml:mfrac>`
	`<mml:mn>1</mml:mn>`
	`<mml:mn>2</mml:mn>`
	`</mml:mfrac>`
	`<mml:mo stretchy='false'>[</mml:mo>`
	`<mml:msup>`
	`<mml:mi>λ</mml:mi>`
	`<mml:mn>2</mml:mn>`
	`</mml:msup>`
	`<mml:mo>−</mml:mo>`
	`<mml:msubsup>`
	`<mml:mrow>`
	`<mml:mo stretchy='false'>(</mml:mo>`
	`<mml:mi>λ</mml:mi>`
	`<mml:mo>−</mml:mo>`
	`<mml:mi>t</mml:mi>`
	`<mml:mo stretchy='false'>)</mml:mo>`
	`</mml:mrow>`

Table continued from previous page.

|	
	```<mml:mo>+</mml:mo> <mml:mn>2</mml:mn> </mml:msubsup> <mml:mo stretchy='false'>]</mml:mo> </mml:math>```
Rendered Equation	$$p_\lambda(t) = \frac{1}{2}[\lambda^2 - (\lambda - t)_+^2]$$

XML in Communication

XML is also used for much of the communication and data handoffs between systems. Incoming deliveries from publishers submitting via FTP delivery include metadata documents in XML which contain information about the article needed to start a manuscript record in the system.

Packages sent to the conversion vendors include metadata documents that contain system-specific information that must be included in the tagged manuscript.

QA of Converted Materials

QA In NIHMS
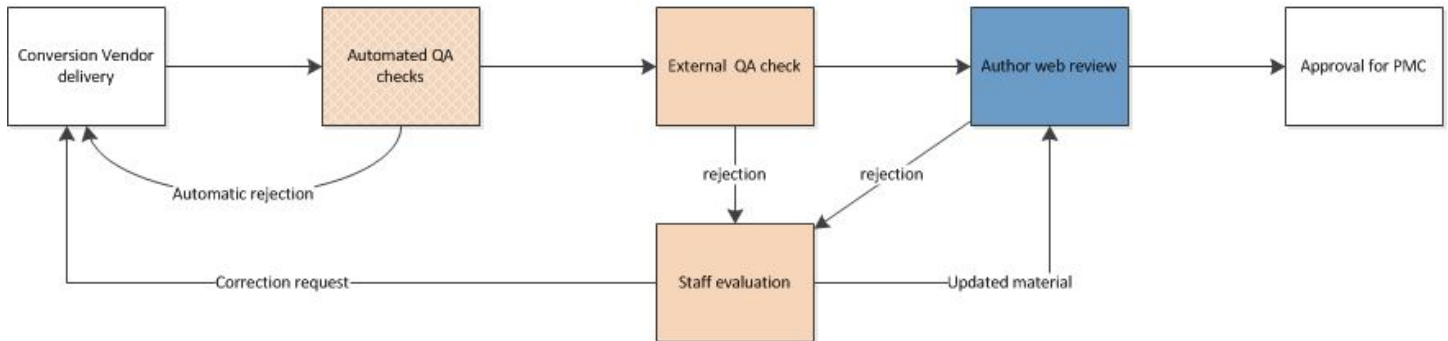
Converted material in NIHMS undergoes several QA steps by various groups before it may be sent to PMC.

The first QA step consists of automated checks for validity and conformity to NIHMS standards of files. Failures at this level are returned directly to the conversion vendors. When the delivery passes these checks, the system generates HTML and PDF versions of the manuscript.

Once these display versions have been generated, the manuscript is sent to an external QA team that confirms that no errors were introduced in either the conversion or rendering processes. Manuscripts that fail this outside evaluation are directed to NIHMS staff for further investigation. Manuscripts that pass this evaluation are sent to the author for review and approval of the converted material.

The assigned reviewing author for the manuscript reviews the display versions (HMTL and PDF) and may either accept the manuscript as is or reject it with comments. Authors may report not only errors in conversion or rendering, but request any necessary updates to the material to ensure scientific accuracy.

Material rejected by either external QA or authors is evaluated by NIHMS staff. Staff may return material to the conversion vendors for updates, request additional files from the submitter or author, make manual updates, or request rendering updates in the NIHMS system.

Additional Processing
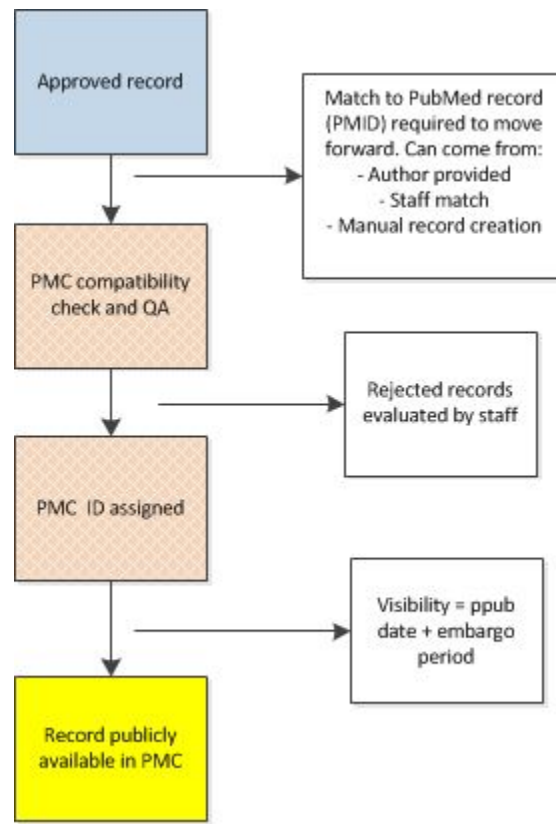
Delivery to PubMed Central and subsequent assignment of a PubMed Central ID number (PMCID) is not automatic after author approval of the NIHMS Web version. The record must also be matched to a PubMed citation to confirm publication and calculate release dates in PMC. This is fairly straight forward for records started from a PMID match by the authors, or deposited with trusted publication information from the publisher that may be used to perform an automated match. For records not matched by these methods, staff must either make a match based on suggested matches by the system, manual PubMed searches, or matches provided by the author separately from the submission of the manuscript.

For records from journals that are not indexed in MEDLINE, a manual citation must be created for the NIHMS record. This involves National Library of Medicine team members reviewing the manuscript record to match the material to information in a variety of external citation indexing sites or publishers' websites to confirm publication information. In rare cases, for journals not indexed in MEDLINE that do not have a publicly-accessible online presence, we may request archive documents from the author in order to verify the citation.

Once the Web version has been approved by the author and matched to a citation record, the record may be sent to PubMed Central. PMC then performs additional automated QA checks on the submission. If the manuscript passes these checks, a PMCID is generated and assigned to the record. If the record fails these checks, it is returned to NIHMS staff for evaluation. Records may also be returned to NIHMS staff following regular PMC integrity checks.

Records in PMC will remain "hidden" until at least the final publication date of the record, plus any additional embargo period specified by the submitter or author.

Reporting Systems And NIHMS
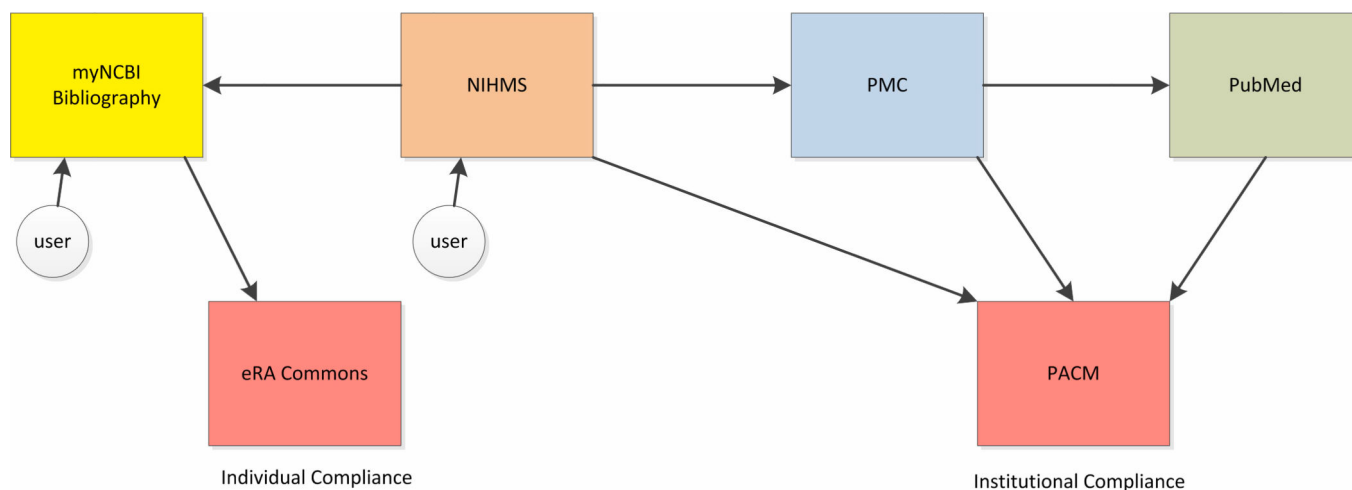
In most cases, records created in NIHMS will generate a PMCID, which demonstrates full compliance with the NIH Public Access Policy, although some records in NIHMS are created only to report the association of a grant with a particular paper. NIHMS IDs are an important aspect of both documenting the relationship between grants and articles and the compliance status of those articles.

The NIH Public Access Policy calls for submission of records for non-PMC journals to NIHMS at the time of manuscript acceptance; there is no grace period in the requirements. Lack of an NIHMS ID for applicable papers will cause authors to be reported as non-compliant. NIHMS IDs may be used by NIH-funded parties as evidence of compliance with the Public Access Policy for up to 90 days after the print publication date of the article, after which a PMCID must be used. An NIHMS ID is not permanent evidence of compliance.

However, NIHMS does not report compliance directly. The system supplies grant and paper associations to a variety of resources and provides status information of NIHMS IDs and PMCIDs when assigned. From these, resources compliance monitoring systems—such as PACM, which provides compliance reports to institutions, and eRA Commons, which provides compliance reports for an individual—calculate article compliance status with respect to the NIH Public Access Policy.