**Ingest Criteria Review**

**Collections & Ingest Committee:** Grace Duke, Christine Giannoni, Bianca Lipscomb, Becky Morin, Matt Person, Suzanne Pilsk, Connie Rinaldo, Judy Warnement, Don Wheeler

**Glossary**

* **Ingest Criteria:**a predetermined list of LC subject headings and call numbers derived from the BHL collection and used to match against materials in the entire IA corpus that have NOT been scanned by BHL member libraries.
* **Ingest:** The bringing in of content from IA that has been contributed by non-BHL member libraries. This process happens on a weekly basis and is based on the Ingest Criteria.
* **Harvest:** The regular process by which all content scanned by BHL member libraries is copied from IA and brought into the BHL database
* **"biodiversity" collection:** Content in IA that is flagged for inclusion into the BHL database. The term "biodiversity" appears in the IA <collection> field. As part of the Harvest, these records are automatically incorporated into the BHL collection.
* **"non-biodiversity" collection:** Content in IA that does NOT have the term "biodiversity" in the IA <collection> field. This content makes up the pool from which the Ingest Criteria is matched to identify relevant content for ingest into the BHL collection.

**Irrelevant Content**

The ongoing ingest of IA “non-biodiversity” collections, i.e. materials brought in as a result of non-BHL member scanning, has resulted in the acquisition of both relevant and irrelevant materials. The BHL Collections Committee has reviewed some of these materials and has identified the following examples of irrelevant content:

|  |  |  |
| --- | --- | --- |
| **Irrelevant Material** | **Suspected Ingest Criterion** | **Recommendation** |
| Materials ingested from the Joint Committee on Taxation ~ 70+ titles | gov't doc nos. in 099 field read as LC call nos. for Agriculture ex. "**S.** Prt. Vol. 674-0007" | Do not match to call nos. in 099 field |
| [Validation of ice skating protocol to predict aerobic power in hockey players / by Nicholas J. Petrella.](http://www.biodiversitylibrary.org/bibliography/36503%22%20%5Ct%20%22_blank) | Respiration as subject heading | "Respiration" irrelevant term, remove from Ingest Criteria |
| [A measure of meter conservation in music, based on Piaget's theory / by Mary Louise Serafine.](http://biodiversitylibrary.org/bibliography/39878%22%20%5Ct%20%22_blank) | "**S**483m" in 099 read as LC call no. for Agriculture | Do not match to call nos. in 099 field |
| [Keeping the body in health, by M. V. O'Shea and J. H. Kellogg](http://www.biodiversitylibrary.org/bibliography/33502%22%20%5Ct%20%22_blank) | Physiology as subject heading | "Physiology" too broad a term, remove from Ingest Criteria |
| [An abridgment of the hygienic physiology : with special reference to alcoholic drinks and narcotics. For the use of junior classes and common schools / by Joel Dorman Steele.](http://www.biodiversitylibrary.org/bibliography/24212%22%20%5Ct%20%22_blank) | Physiology as subject heading | "Physiology" too broad a term, remove from Ingest Criteria |

There is a need to review the Ingest Criteria, a series of LCSH terms and call nos. used to match against the IA corpus, and refine the list to include only those terms and call nos. that yield the highest return of relevant content. The Committee has implemented 4 strategies for review:

**Review BHL Terms**

It is necessary to revisit the list of non-geographic BHL subject keywords (derived from the MARC ‘650 a’) and identify terms that are specifically relevant to the various disciplines that contribute to the study of biodiversity. While there are many subject headings that are relevant to biodiversity, some are much too broad in scope and do provide adequate ROI when used as part of the Ingest Criteria. The following terms are under consideration for removal from the Ingest Criteria list:

* Physiology
* Reproduction
* Genetics
* Anatomy

**Identify Good Ingest Terms**

What subject terms are coming in as a result of the Ingest completed to date that are NOT already BHL subject terms and would be good additions to the Ingest Criteria? With the continuing ingest of "non-biodiversity collection" materials from IA, we are bringing in new subject headings to our database that may be unique to the BHL collection. Looking more closely at these terms, it may be possible to identify relevant terms that could be added to the ingest criteria. In this way, the IA Ingest is "learning" from its own collection of subject headings unique to the BHL collection. “Wilderness areas” and “Wildlife management” for example.

**Identify Bad Ingest Terms**

What subject terms are coming in as a result of the Ingest completed to date that are absolutely irrelevant to the BHL collection and should be marked for exclusion as part of the Ingest Criteria? This is an ALL OR NOTHING approach. Any subject terms identified in this category would override any other terms also associated with the record and negate the record. For example: BHL should exclude all records with the LCSH term = “God” from the Ingest pool regardless of any other terms like "Birds" or "Darwin" also associated with the records. If the term "God" is used as part of the MARC '650 a' that means the work has a great deal to do with the subject.

**Classification/Call Number Review**

It is necessary to refine the list of LC call numbers used as part of the Ingest Criteria.  Currently, the Ingest Criteria takes a broad approach by matching at selected LC class ("Q" -- General Science) and subclass ("QH" -- Botany) levels.  Mike L. has described that while "tricky" there may be the opportunity to refine the criteria further to include specific number ranges, such as "QE 700--999" for Paleontology, a subset of the "QE" subclass for Geology.  LC classes/subclasses that are evaluated as being too broad in scope will either be refined or eliminated altogether if further refinement is not easily executable.

Dewey numbers are under consideration for inclusion into the Ingest Criteria.  As Dewey numbers allow for finer grained classification, they may prove to be useful additions.  Dewey nos. will be selected to match against the LC classes/subclasses selected.

Finally, it has been decided that only the standard Dewey and LC classification MARC fields will be targeted for matching against the Ingest Criteria.  Standard LC class/call no. MARC fields are the 050 and 090 fields.  Standard Dewey MARC fields are the 082 and 092 fields.  Other non-standard fields that may hold class/call nos. such as the 099 and 852 fields will NOT be targeted for matching, as these fields have proved insufficient for ingesting relevant content.

**For further questions or concerns, please contact Bianca Lipscomb, lipscombb@si.edu | 202-633-2239**