David Ruddy
Cornell University Library
Project Euclid and arXiv.org

Project Euclid (http://projecteuclid.org)

1. Content: size, diversity, multilingualism, digitized vs. born digital, rights (public domain vs. copyrighted), access status:

    Current size: over 110,000 journal articles from 65 math journals.
    162 monographs and 1 large conference proceeding.
    Continuous growth, both in new titles and current content of existing titles.

    Diversity: journal, conference proceedings, and monographic research literature in mathematics and statistics. Most content published since 1940, but some titles are older (oldest journal issues are from 1891; monographs from the 19th century).

    Language: mostly English, with a moderate amount of French and German content.

    Born digital: Euclid includes a substantial number of post-1995, TeX produced PDFs; yet the majority of pages currently in the system are digitized from paper.

    Rights: copyright remains with publishers. Some content is in the public domain.

    Access status: both open and subscription controlled, as determined by the content's publisher. Currently, 70% of material in Euclid is freely accessible.

2. Business model: a not-for-profit, academically owned and controlled initiative in publishing, jointly operated by Cornell University Library and Duke University Press. Euclid operates on a cost-recovery model (revenues from hosting fees and subscribed content support operations and open access content).

3. Copyright: most content is in copyright, held by publisher. Some public domain content.

4. Access policy: varies, as publisher determines. Euclid itself has established a 5-year moving wall (older than 5 years is open) on the one bundle of journal content that it sells to libraries (Euclid Prime: approximately 30 journal titles).

5. Preservation: the responsibility of Cornell University Library.

6. Interoperability: Euclid shares metadata via OAI-PMH and RSS; shares content with discovery services (Google, Microsoft, Serials Solutions, ExLibris, etc.); metadata deposited with CrossRef.

7. Unique features: MathJax; seamless mix of formats (journals, books, etc.); reference linking.

8. Math-aware features: Euclid is a disciplinary repository—its holdings are in math and statistics only; ability to search on MSC codes, or portions of codes; links to MR and ZBL.

9. Planned extensions/enhancements: currently working on a new user interface with additional user features.

10. Expectations from a global DML effort: reasonable, low-barrier methods for participating in global math network; shared math unique tools.

The arXiv math collection (http://arXiv.org/archive/math)

1. Content: size, diversity, multilingualism, digitized vs. born digital, rights (public domain vs. copyrighted), access status:

    Current size: over 150,000 articles with math.XX as primary classification.
    Another 17,000 articles cross listed to a math category.

    Diversity: almost exclusively current work, dating from 1992. Math arXiv represents approximately 18% of total arXiv.org holdings (1991-current). Math submissions in 2011 were about 26% of total arXiv submissions.

    Language: almost all in English; titles and abstracts must be in English.

    Born digital: yes.

    Rights: copyright remains with author.

    Access status: open.

2. Business model: as of recently, arXiv is community supported (heaviest institutional users) via a membership model.

3. Copyright: arXiv has a perpetual, non-exclusive right to distribute submitted articles. All other rights are left with the author.

4. Access policy: arXiv content is freely accessible to anyone.

5. Preservation: the responsibility of Cornell University Library.

6. Interoperability: OAI-PMH metadata in DC and other formats; RSS feeds; arXiv API (real-time programmatic access to metadata and search engine); bulk downloads available; ten mirrors worldwide; SWORD deposit interface and other data feed channels for content collection.

7. Unique features: author submissions rather than publisher content (although some publishers submit content); strictly enforced version control; daily additions to content; extremely active, worldwide use; alternate interfaces into arXiv allowed (e.g., http://front.math.ucdavis.edu).

8. Math-aware features: submissions are classified into math subject areas; authors can include MSC codes as part of submission metadata.

9. Planned extensions/enhancements: arXiv is currently implementing Invenio as its front end.

10. Expectations from a global DML effort: reasonable, low-barrier methods for participating in global math network.