

# The Deposit Policy: Balancing Content Goals and Ingest Control

By

David B. Lowe, Digital Collections Management Librarian, Texas A&M University Libraries

Charity K.M. Stokes, Cataloging and Metadata Librarian, Texas A&M University Libraries

## I. The institution's IP and OA context

There are a host of implications and risks associated with the act of placing any author's writing in the public sphere. As in the case of an institutional repository (IR), the mere fact of a university declaring: "Here are works related to this institution" is a bold act that opens up vulnerabilities just as it allows that organization to establish its identity, demonstrate its importance and quality, and publicize its impactful contributions to society. What is a university to do?

A comprehensive deposit policy for the local IR is the essential tool to address the risks, but it should align first of all with the institution's intellectual property (IP) policy and any open access (OA) policies relevant to campus authors, and secondly, it should feature a local governance body or individual empowered to adjudicate, amend it, or **waive requirements** in special circumstances.

In the spring of 2019, the authors conducted a review of policies relevant to Association of Research Libraries (ARL) institutions, verifying that virtually all ARLs are part of universities with IP policies acknowledging that copyright in scholarly publications belongs to the individual authors where that university's affiliates are concerned. (The gray area tends to be instructional materials, which are often made available in IRs; some universities do treat this teaching space as essentially a "work for hire" condition, asserting that copyright for these items belongs to the employing institution.) The distinction is critical because if authors do not hold the copyright, they are mostly removed from the decision-making process where deposit is concerned. This chapter will focus primarily on the context of an academic institution where campus authors (whether students or faculty or other employees) retain the copyright in their original creative and scholarly works that are under consideration for deposit in an IR. To locate the policy at a given institution, the most effective method might be to search for "intellectual property policy" or "copyright policy" on that institution's website. ~~For reference, data gathered by the authors may be found at: [doi.org/10.21203/rs.3.rs-1111111/v1](#).~~

Of similar importance, the presence of an OA deposit mandate at an institution often involves the IR. For reference, it may be helpful to consult the Registry of Open Access Repository Mandates and Policies (ROARMAP: <http://roarmap.eprints.org/>). In some cases, even when a campus lacks an OA policy, still some units within the larger organization may have adopted such a policy, and ROARMAP is an attempt to record the details of all these cases. Also worth keeping in mind is the fact that not every policy requires deposit in the *local* IR, but when it does, the IR deposit policy will need to be as supportive as possible of the OA stipulations, including such factors as

43 timeframe requirements. Any bureaucratic barriers from the IR side would be  
44 counterproductive toward the OA cause, so it is in the interest of IR managers to make  
45 OA policy compliance as easy as possible for all involved.

46 These broader, more philosophical concerns can vary among universities, but they  
47 define the playing field and direct the more practical, quotidian workflows that this  
48 chapter will navigate.

## 49 II. Who: Authors and Depositors

50 One of the most basic policy components for any IR is stipulating who will be allowed  
51 to have their materials hosted. A second set of actors to be defined is who performs the  
52 work of depositing. Related policy needs include what types of items may be added  
53 into the repository.

← p. 52

p. 53 →

54 As an instrument of the institution and a gauge of its significance, in common practice  
55 the IR is generally open for deposit only to affiliates to share scholarly, research  
56 materials. Specifications regarding which categories of affiliates can deposit and who  
57 actually does the depositing can vary widely from institution to institution and may  
58 depend on local factors, such as the scale of the procedural support required and the  
59 related staffing available.

60 With decreasing levels of consensus, here are the main groups of affiliates that might  
61 be considered relevant depositors to an IR:

- 62 1) Faculty
- 63 2) The library itself
- 64 3) Research staff
- 65 4) Graduate students
- 66 5) Undergraduate students
- 67 6) Alumni
- 68 7) Community members

69 In the early literature on building IRs, faculty are unquestionably the most anticipated  
70 group to share their scholarly materials in the IR. As discussed above, depending on  
71 the wording of a campus OA policy, they may even be **required** to deposit their  
72 publications. Secondly, libraries may generate much of the content in an IR (which  
73 some places may consider as separate “digital collections” territory and not that of an  
74 IR), either through digitization workflows or by routing in publications by the  
75 university, like yearbooks and newspapers, or such born-digital institutional records  
76 as Board of Trustees minutes. To the extent that librarians may not be considered  
77 faculty on some campuses, here again their individual scholarly efforts are generally  
78 considered copacetic inclusions. Non-faculty research staff would likewise be prime  
79 candidates for sharing relevant works.

80 Moving on to the next layer of submissions, graduate students’ theses and  
81 dissertations have become one of the largest sets of document types for most IRs, but  
82 from here down on this list, the waters become murkier. Anecdotally, while there are  
83 faculty who would welcome wider inclusion of students’ work, still there have been

84 professors who bemoan the inclusion of even certain graduate theses that they  
85 consider unworthy. It is important to keep in mind these perceptions of quality and  
86 reputation to strike a balance with the policy. Without question, graduate student  
87 articles in professional journals would also be deserving of a home in the IR, so  
88 deposit of such materials should be given consideration, including an ingest  
89 procedure.

90 For undergraduates, the main set of consensus materials for the IR includes honors  
91 theses, capstones, and portfolio-style sets of materials that represent significant  
92 achievements. Beyond that, there may be an occasional undergraduate literary  
93 periodical in the IR at some institutions, but **the quality** marker for what might be p. 54 →  
94 considered “scholarly” tends to end about there for this level. Although alumni and  
95 even perhaps community researchers may seek to add their works, it would be rare for  
96 most IR administrators to consider such materials without a faculty collaboration or  
97 similar active connection to the institution. In such cases, it would be advisable to  
98 draft a Memorandum of Understanding (MOU) to establish expectations and norms to  
99 scope the extent of the shared work and preempt mission creep.

100 Having scoped whose materials might be a fit for the IR, the next vital decision would  
101 be determining who handles the items’ deposit process. Allowing self-deposit for  
102 faculty may appeal greatly to some: with author-supplied documents and  
103 descriptions, the library gains content without diverting staff to cover those tasks.  
104 And—the thinking goes—who better to describe an item than the person who created  
105 it? By means of a simple form, structured metadata would be simple to produce,  
106 making the items easily findable. Certainly any place that is too understaffed to  
107 mediate OA submissions will have little choice but to follow the unmediated path.  
108 Those that can mediate submissions will find plenty of work in navigating pre-print,  
109 post-print, and publisher’s versions of papers and in otherwise correcting and  
110 enhancing metadata.

111 However, if scale presents a formidable barrier for the ingest process, it becomes an  
112 even greater one on the discovery side over time. The challenge of growing digital  
113 collections is not a simple one; as collections grow, so does their internal complexity.  
114 Different subjects require different handling and labeling of those subjects. Different  
115 formats require different types of metadata. For example, basic fields are good for  
116 textual items such as books or articles, but not so good for archival items such as  
117 letters or photographs (which often do not have titles). These differing types of formats  
118 often require specialized metadata fields in order to facilitate discoverability. Add to  
119 this complexity the fact that different contributors may enter names in a variety of  
120 ways, and the difficulties begin to mount in some very tangible ways. For example, in  
121 the case of finding all works in a collection authored by or otherwise connected to the  
122 same person, it soon becomes obvious that typographical errors and misspellings are  
123 common, which hinders comprehensive discoverability. And the list of complications  
124 does not end there; for example, system handling of special characters, name changes,  
125 title changes, version control—all can combine to create a challenging environment.  
126 Because of these issues, it is necessary to have clear policies and procedures on how  
127 to enter metadata into a collection. These policies need to include the **following:** ← p. 54

- 128 1. **Clear** definitions of expectations for each metadata element
- 129 2. Instructions on how to input certain information, such as how to structure
- 130 names and dates
- 131 3. Recommendations on quality review of metadata before ingest to a production
- 132 server

133 Being flexible with staffing arrangements helps to accomplish all of this IR work in the  
134 context of resources, which can vary widely across institutions.

### 135 **III. What: Scholarly Materials**

136 Earlier this chapter has covered questions of who might deposit, but the content  
137 decisions are closely connected to the roles of these author groupings. Stipulating  
138 what materials are suitable for the IR is akin to a collection development policy. The  
139 major differences may be expressed as follows: traditional library collection  
140 development is more goal-driven and anticipatory of what research may be produced  
141 in the future, while decisions about IR collections are more often than not reactive,  
142 reflecting the authors' prior decisions about what research to produce and then share.

143 Having stated what groups may deposit in the IR, the scope of what may be deposited  
144 narrows. First of all, supporting OA, whether mandated on a particular campus or  
145 not, is one of the mission critical reasons for libraries to even have IRs. Deposit policy  
146 decision makers should be open to allowing not only scholarly articles, but the related  
147 documents in that life cycle, like conference presentations or the foundational  
148 research data (whether tabular data, or sets of images or audio or video, etc.) upon  
149 which the papers are based. Journal editors on campus may find hosting of their  
150 scholarly journals to be of interest, especially if the platform is accommodating of the  
151 editorial process and pleasingly professional in its presentation of the material. Also  
152 not uncommon for consideration for deposit from professors would be their  
153 instructional material if, as discussed above, copyright is not an issue. The growing  
154 success of open educational resources (OER) proves that this curricular support path  
155 is worthy of consideration for IRs, with the added marketing feature of reducing costs  
156 for students. However, the leading efforts, such as OpenStax.org, have developed their  
157 own distribution platforms, so OERs may not require local IR deposit to be successful.

158 For students, both graduate and undergraduate alike, the typical document submitted  
159 would be a thesis or dissertation, in many ways similar **in format** to the faculty textual  
160 documents. Presentations are commonly sets of slides, while tabular data sets are  
161 usually manifested as spreadsheets. Still images, audio, and video round out the top  
162 format types for which authors seek deposit in an IR.

163 For all format specification decisions, it may be useful to consult the Federal Agencies  
164 Digitization Guidelines Initiative (FADGI) recommendations  
165 (<http://www.digitizationguidelines.gov/guidelines/digitize-technical.html>), in addition to  
166 conducting environmental scans of peer institutions. The clarion call central to the  
167 long-term preservation function of IRs is to prefer open formats that will not be  
168 dependent on proprietary software.

169 Since scholarly textual documents are commonly authored in word-processing  
170 software like Microsoft Word or LaTeX, then converted to PDF, the act of conversion to  
171 the destination format can sometimes introduce unintended consequences related to  
172 page formatting. Add to this complication the fact that the more preservation-minded  
173 IR programs require the archival format known as PDF/A (of which there are yet again  
174 more layers of levels and versions). PDF/A represents an effort to make each file self-  
175 sufficient, without dependencies on particular external fonts, for example, that may  
176 not make it through all the transitions the document may take over time, like  
177 migrations across hosting servers or version changes in display software. A further  
178 twist to the format story is that the accessibility needs of the visually impaired have  
179 not been adequately addressed throughout the history of IRs, and formats such as  
180 PDF/UA are deserving of increased consideration. The challenge will be to find a  
181 happy medium between accessibility and preservation, ideally a file that is both  
182 PDF/A and PDF/UA compliant. Efforts are sure to be ongoing with these issues into  
183 the third decade of the 21<sup>st</sup> Century. A forward-thinking policy will incorporate and  
184 balance both the accessibility as well as the digital preservation requirements  
185 advocated within the institution and beyond.

186 An important piece of documentation and part of the submission process for all of  
187 these formats not yet discussed here are the author agreements. Functionally, the  
188 essential components of such documentation are:

- 189 1. to ask depositors to confirm that they are the authors
- 190 2. to verify that they have not included the copyrighted work of others without  
191 giving credit
- 192 3. to grant the institution a non-exclusive license to make the material accessible.

193 Even this simple licensing procedure often varies across the collection types. It may  
194 range from paper forms to electronic documents to a streamlined (and too often taken  
195 for granted) click-through **agreement**. **Deposit** procedures can often take place below  
196 the surface in IR documentation, so environmental scans—the go-to technique for  
197 policy development in libraries—may not reveal much in the way of options, so a  
198 conversation with IR managers might be a helpful exercise in formulating this policy  
199 component.

#### 200 **IV. When: Academic Life Cycles**

201 It would not be uncommon for electronic theses and dissertations (ETDs) to comprise the bulk  
202 of materials in an IR. Since the vast majority of IRs are in academic institutions, there is a  
203 tendency for deposited items to follow cycles related to the academic year. Theses and  
204 dissertations appear in bulk typically from April to June, with a lesser amount being added  
205 before winter break and some at the end of summer. But other materials, such as conference  
206 presentations and proceedings can appear when faculty are preparing for or returning from  
207 meetings; other materials, such as publications added by centers associated with the institution,  
208 can likewise follow predictable rhythms.

209 Beyond these academically cycled types, other publications will appear more sporadically. It is  
210 very easy to lose track of when these publications need to be added, especially if the deposit



211 workflow is dependent on library resources. A policy or timetable listing these times of year,  
212 along with what materials are expected, goes a long way towards helping IR staff keep on top  
213 of adding materials in a timely manner. It is also recommended that there be a commitment to  
214 those whose items are being added to the repository so that they know when they can expect  
215 their materials to be available. With ETDs, this may be handled by the appropriate graduate  
216 office dealing with the theses and dissertations. For other materials, letting contributors know  
217 that their items will be available through the IR in a designated amount of time (two weeks, a  
218 month, etc.) will not only provide a way to keep work from piling up, but also give staff time to  
219 process materials without becoming overwhelmed. Beyond policy, for ongoing collections,  
220 good communication between the originating departments and those who may assist with  
221 ingest is essential, since staff turnover and shifting priorities can lead to a host of orphaned  
222 collections.

## 223 **V. Where: Managing Platforms**

224 Some institutions only have one platform for their IR, while others have more than one that  
225 they use for particular types, treating each platform separately. It is a good idea to have a  
226 cohesive policy regarding what material will be hosted on which platform, based on what  
227 provides the user the best experience accessing the materials.

228 An institution-wide policy covering the various types of platforms and treating all materials  
229 and related discovery tools as part of the overarching repository will facilitate coordination of  
230 this highly complex endeavor. Coordination may require concessions among various  
231 departments who have a stake in their particular platform, and may even require  
232 administrative shifts to facilitate that coordination. As is frequently the case with institutional  
233 change, it may not be easy or comfortable. However, in order to regard collections as part of the  
234 overarching information ecosystem managed by the campus library, consistency among the  
235 various platforms is key to an optimized user experience. There is a significant intersection here  
236 with discovery services. But having a policy that scopes out the mission and reach of the IR will  
237 provide users more seamless service.

## 238 **VI. How: Procedural Scenarios**

239 Although it may be inefficient to do so, the mechanics of ingesting one item at a time are always  
240 an option with any repository platform. To meet the scale of the challenge that larger sets  
241 present, however, IR staff will need a solution for batch ingests. Some institutions add high  
242 volume materials like ETDs programmatically, using software that conforms to the Simple  
243 Web-Service Offering Repository Deposit (SWORD) protocol, linking established approval  
244 workflows to smoothly add metadata, licensing documentation, and the items themselves to the  
245 IR. Vireo, from the Texas Digital Library, is one such widely used solution. Alternatively, other  
246 institutions may rely on metadata librarians and IT staff to do bulk ingests to the IR. In any  
247 case, the connection with the deposit policy would include the need to be realistic about  
248 capacity and satisfying expectations where timeframes are concerned.

249 Another timeframe aspect integral to ETDs is the question of embargoes. While the vast  
250 majority of ETD authors do make their works available immediately, still the option to withhold

251 a work from public access is an option that most IR managers will need to allow. The two main  
252 reasons for choosing embargoes are related to future, pending publications, including journal  
253 articles and monographs, but also patent processing. Embargoes tend to be favored in specific  
254 disciplines like engineering and related innovation-dependent fields on the one hand, and on  
255 the other, humanities fields like English and history, where authors depend on a dissertation-  
256 based monograph to be a commercially viable publication. IR policies need to account for this  
257 need, while admittedly librarians will not be too shy to encourage open access as advocacy  
258 venues avail themselves. Functionally then, the policy will need to account for setting,  
259 extending, and removing embargoes. In practice, a typical institution may allow the initial  
260 embargo period to extend for six months to two years. Most allow the embargo to be extended  
261 or renewed in time units up to two years.

262 If it is important to delay release for some things, it is also reasonable to account for completely  
263 removing others. Especially when deposit is not mediated, there will always be the risk that  
264 some material gets posted that will need to be taken down, whether due to a copyright  
265 infringement, retraction, or some other regrettable circumstance. Thus, the take-down  
266 procedure as an element of repository policy rears its ugly head. Such a policy needs a  
267 reporting procedure, such as a form. That step may serve as an opportunity to state clearly that  
268 the collections the library makes available are either in the public domain or licensed to be open  
269 by their authors. The form should gather information about the allegedly infringing work on  
270 the repository site as well as contact information from the copyright holders or their  
271 representatives. Finally, it should indicate a reasonable time period for resolving the issue. As  
272 mentioned briefly above, there must be an individual or body charged with adjudicating and  
273 amending the policy and waiving certain requirements when circumstances dictate, and take-  
274 down cases are exhibit #1.

275 **VII. ~~Case studies~~ Sticky Scenarios**

276 No amount of policy or licensing shielding crafted with the best of legalese can prevent sticky  
277 situations. Privacy issues, for example, can still creep in. In one case familiar to the authors, an  
278 alumnus of a large public university was seeking to escape the all-seeing eye of search engines.  
279 Although he authored no material in the IR, he petitioned to have his name removed from a  
280 commencement program that was included among institutional archival materials on the site.  
281 Ultimately, as an institutional record of a public entity, the graduation list was deemed by the  
282 IR admins to be a document that should not be altered. Increasing scrutiny of IRs in the wake  
283 of such specific phenomena as the European Union's General Data Protection Regulation  
284 (GDPR) or more generally movements like the "right to be forgotten" will doubtlessly make  
285 more cases like this one more of a concern in the future.

286 Another privacy case of note involved the honors thesis of a recent undergraduate. The subject  
287 matter of the paper was such that she feared it would be the subject of concern in her career as a  
288 public school teacher. In that case, the IR administrators agreed to suppress that particular  
289 work.

290 More well-known from popular and social media at the time of this writing might be the case of  
291 embarrassing news or regrettable photographs in campus publications such as newspapers and  
292 yearbooks. IR administrators are put in the particularly difficult position of anticipating  
293 precisely what of the historical record might come back to haunt us later, but integrity lies in an  
294 unswerving commitment to transparency and the authenticity of the documents that have  
295 arrived in a librarian's hands through some established chain of custody.

296 One final case study that will not be surprising on campuses with strong STEM programs  
297 involves tools and methods for technologies that have military applications. In one case  
298 familiar to the authors, a dissertation related to rocket engine optimization was removed from  
299 campus in all of its manifestations, including all paper copies. While it is not the IR manager's  
300 function to perform this role of national security gatekeeper, it is important to know where the  
301 boundaries may be for openly accessible scholarly materials. Since openness is a common goal  
302 in scholarly communication, there may be opportunities to advise budding scholars that the  
303 choice of certain topics may preclude participating in expanded discourse about their  
304 specialties, so the earlier they are aware of the implications of their choices, the better.

305 In conclusion, some have compared open source software solutions as being free like a free  
306 puppy. And so it is with IRs. The challenge of maintaining such an entity requires extensive  
307 commitments that relate to necessary policy elements. This chapter has summarized some of  
308 the top policy elements that can set an IR up to be successful.