Break, Drift, Rot: How Academic Librarians Can Weatherproof References in Electronic Theses and Dissertations

Kathy Anders, Sarah Potvin, Tina Budzise-Weaver, Martin Klein
Who We Are

Tina Budzise-Weaver
Associate Professor
School of Performance, Visualization & Fine Arts
TAMU

Sarah Potvin
Associate Professor
Department of English
TAMU

Martin Klein
Computer Scientist
Los Alamos National Laboratory

Kathy Anders
Associate Professor
Department of English
TAMU
Overview:

1. Links in ETDs rot and drift
2. We are trying to figure out how to make that happen less
But first some background...
Reference rot is a term we introduced in the Hiberlink project to denote the combination of two problems involved in using URI references, both of which relate to the dynamic and ephemeral nature of the web:

**Link rot:** The resource identified by a URI may cease to exist and hence a URI reference to that resource will no longer provide access to referenced content.

**Content drift:** The resource identified by a URI may change over time and hence, the content at the end of the URI may evolve, even to such an extent that it ceases to be representative of the content that was originally referenced.


Put more simply -

**Link Rot:** Link stops working e.g., HTTP 404

**Content Drift:** Linked content changes over time

**Reference Rot = Link Rot + Content Drift**
Background

What do we mean by rot?

https://thetellingproject.org/iowa-review-publication/
Background

What do we mean by drift?
Is this a problem in ETDs? (Yes!)

Analysis of 484 links in TAMU Performance Studies from 2012 to 2020

Link Resolution and Functionality

- Perfectly Functional: 251 (51.9%)
- Not Resolved: 128 (26.4%)
- Diminished: 105 (21.7%)
What are some of the concerns?

1. Scholars will lose records of and access to their sources and data
2. In some cases there may be no other copies of materials, particularly those related to sources from marginalized, underrepresented, or disempowered groups
So what do we do?
(Socio-technical Interventions)
Coordinate with Stakeholders

- Work with systems developers
- Talk to ETD offices
- Reach out to graduate advisors/coordinators
- Outreach to librarians
Educate Students

- Workshops
- Graduate classes
- Guides
- Point-of-need
What can students do?

- Make robust links
- Describe key information in text
- Consider changes to citation styles to capture more information
- Think about ethical implications about how information might be shared (including the right to be forgotten)
But we know that ultimately people can do only so much...
Technical Interventions

- Invest in web archiving programs like PermaCC for students to use
- Embed web archiving options and prompts into systems like Vireo
- Create large scale batches of robust links as part of the back end of ETD processing
Web Archiving Services and Tools

Create a snapshot of referenced resources in a public web archive

https://robustlinks.mementoweb.org/

http://web.archive.org/
But you can’t preserve a link that isn’t there...
This is why we need socio-technical interventions

We need to educate students and stakeholders while developing technical systems to help streamline and expedite the archiving of web-based sources in ETDs.
Why should librarians do anything about ETD reference rot?

“We collectively bear greater responsibility for this body of scholarly work, and need to move forward from a position of benign neglect to one of informed curation and proactive preservation of an important collection of scholarly output which is at risk.”

What did we do?

- We sent out a survey to all the grad students at TAMU to find out how worried they are about reference rot.
- We offered two workshops to graduate students on this topic (and fed them).
We are processing the data

- The survey had 294 responses
- The workshops had ~25 attendees
- It’s hard to get grad students come to workshops
Thank you

1. Questions?
2. Comments?
3. Questions that are more of a comment than a question?