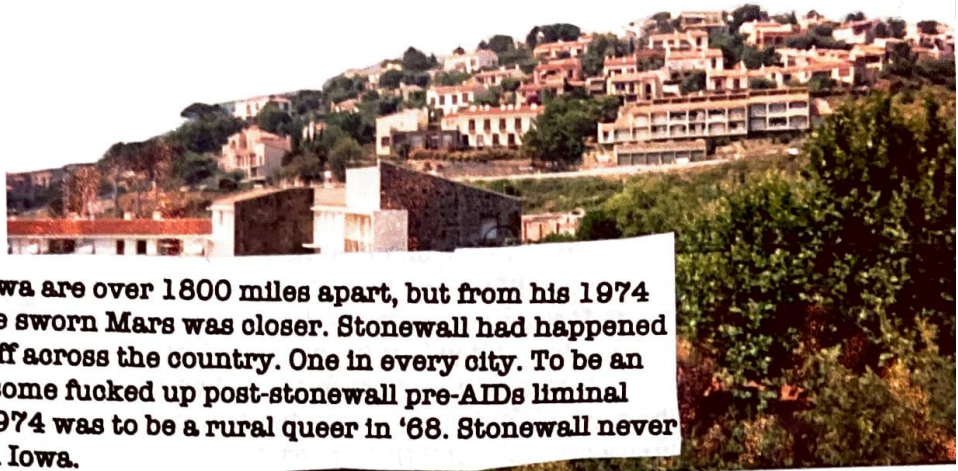




Rather Flawed
Digitization
By Elizabeth Koning &
Elizabeth Schwartz



San Francisco and Grinnell Iowa are over 1800 miles apart, but from his 1974 Grinnell farmhouse, Stewart could've sworn Mars was closer. Stonewall had happened 'bout 5 years ago. Gay lib cells took off across the country. One in every city. To be an urban queer in 1974 was to exist in some fucked up post-stonewall pre-AIDS liminal discotheque. To be a rural queer in 1974 was to be a rural queer in '68. Stonewall never made it to Stewart's place in Grinnell Iowa.



led to success at the uprising--disruption by physically occupying public space as out, loud, proud, defiant queers. This is what we call visibility politics. It's what worked to win the voting rights act in '68, and it's what would work to win the ADA in 1990.

You see, the gay liberation movement was totally urban centric in its tactics. Gay lib is usually understood as beginning in 1969 in Greenwich Village at the Stonewall Inn on Christopher Street. Don't get me wrong--queers had been resisting heterosexism long before stonewall (check out the Compton's Cafeteria uprising), but this uprising did lead to the creation of a national gay liberation front. The Stonewall Inn Uprising went down in history because it just so happened that some white gay guys were there and wrote about the story as if it were their own. And of course the media picked it up, and before ya know it gay lib started at Stonewall. These white gay guys didn't instigate the rebellion--trans women of color and sex workers did that.

I could go on about what really happened at Stonewall for the rest of my pseudo-academic career, but for this story what actually happened that night on Christopher street doesn't matter. What matters is what happened after. Because after Stonewall gay liberationist fronts spread across cities in the US like wild fire. They adopted the strategies that

COMPTON'S CAFETERIA REVOLT - 1966

one August night in the Tenderloin trans women of color, sex workers, & drag queens fought the SF PD's attempts to arrest them and WON!

check out Susan Stryker's Screaming Queen for more on the revolt



But here's the thing with visibility politics: they only work if you've got enough people being visible. They work damn well in places with a big population like New York City. They're a lot less effective in Grinnell Iowa: population 8,402. And so as queers in San Francisco, New York, Chicago, and even Iowa City came out of their closets and into the streets, Stewart was out in his field. He wasn't alone. No, he lived in a farmhouse with some other queers. But even in the company of friends and lovers, it got lonely. The urban gay lib movement--in all its great successes--had left rural queers like Stewart behind.



Then one day Stewart and his comrades packed up their truck and drove to Ann Arbor, Michigan for a gay conference. While there, they met rural queers from all over. Men who loved men and the serene quiet of rural life. They met one man named Stevens McClave from rural Oregon. Stevens had left Oregon for San Francisco a few years back, but he was a country person at heart. He loved show tunes, old rock n' roll, glitter, and his cats. Over a distance of 2,000 miles, the Grinnell collective and Stevens wrote to each other and came up with the idea for a magazine for gay rural people to create a gay lib movement for rural gays too. Before he could see that dream realized, Stevens killed himself.



GONE BUT NOT FORGOTTEN

Stevens' death set a fire under the Grinnell's collective ass. Rural gays were scared, in pain, and alone in the face of deadly heteropatriarchy. The magazine was a lifeline. After months of correspondence among rural gay collectives across the country, weeks spent crouched over a typewriter, and sleepless nights cutting and pasting it all together, it was done:



"Here it is. No longer a fantasy, R.F.D. exists. As gay people living in the country we felt a need for things that urban-oriented gay publications and adamantly heterosexual country magazines could not provide... finally this first issue—Rustic Fairy Dreams—has come true. We hope to break down the feeling of isolation from things gay that many of us experience in rural settings, to build some sense of community among rural gay people, and to provide the means of sharing with each other our thoughts, feelings, and ideas about our unique experience as gay country people. R.F.D. is a reader-participatory venture. You write, sing, dance, and are R.F.D. We need your contributions of material, energy and love to survive" (Allan, Carl, Don-Tevel, Olaf, Ollie, Rick and Stewart 3).

That was from the first page of the first issue dated Autumn 1974. 48 years later, RFD is still a reader-participatory magazine for queer country folks. I ran across RFD when some friends and I stumbled into a queer St. Louis coffee shop on June 7, 2019. At that point in my life, I hadn't fully come out to myself yet—although the fact that I felt so at home in that queer as fuck coffee shop probably should've been a clue... hindsight's always 2020. I'm not a gay man, and I didn't grow up in rural America, but reading through those few issues of RFD made me feel seen in ways I didn't know was possible. There was something about holding a piece of queer history—a history my Catholic-schooling told me didn't exist—that was transformational.

My gut reaction to reading the two issues we found in the coffeshop was: we need to digitize this. This must be preserved and made accessible. Three years later, I feel similarly. Between 2019 and 2022, the current editors of RFD scanned all the back issues and made these scans accessible on issuu.com.

However, these scans are just page images. They remain unsearchable, unindexable, and undiscoverable to anyone who doesn't know exactly what they're looking for. An OCR version of the nearly 50-year-old mag could change that.

So, when given free reign to do whatever I wanted for an experiment, I turned to Elizabeth and asked her if she'd ever heard of RFD and if she wanted to try to digitize it. We spent the first few of our meetings totally engrossed with the magazine. Finally we got around to actually contacting the current editors, researching its history, and experimenting with Tesseract on the issuu.com scans. And here it is. No longer a fantasy, our RFD-experiment exists. As queer students of the digital inclination, we felt a need to digitize this old magazine. We questioned that impulse, and here's what we've got to show for it.


```
(base) queers@queers-Macbook-air ~ % cd OCR_studies  
(base) queers@queers-Macbook-air OCR_studies %
```

Ever since the rise of the field, digital humanists have been studying OCR. OCR produces the searchable texts behind most text-based DH projects and databases (like JSTOR, chronicling america, google books, etc.). However, the software behind OCR was trained to read mid 20th century corporate documents into computerized text. More recent softwares have been trained on 19th British print. As a result, OCR is excellent at accurately transcribing 19th century anglophone texts and corporate documents. For any other kind of project, programmers have to train an existing OCR software on a new corpus to improve its accuracy (Alpert Abrams). Even when the hand-transcribed training data exists, this is a massive undertaking. And because digitization projects that rely on hand transcription are expensive and lengthy, these projects tend to be few and favor canonical white, Anglophone texts, leaving out corpora of non-canonical work in languages other than English and alphabetic (an non-alphabetic) systems that aren't the latin alphabet (Smith and Cordell).

The linguistic hegemony of English in OCR software is a problem that new training corpora and increased funding can address (Smith and Cordell). However, scholars haven't started to question how OCR's visibility politics in and of themselves might be ill suited to digitize queer content. OCR as a software also engenders hegemonies that we haven't bothered to question and digitization of RFD forces us to contend with.

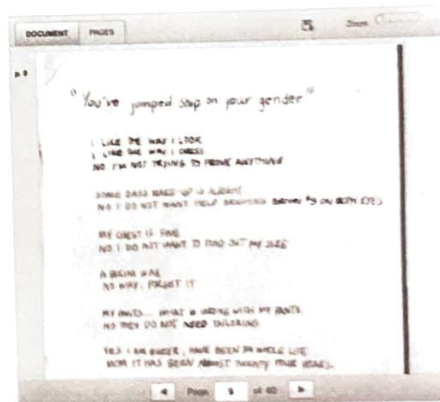


STUDIES

Librarians who work with zines like RFD take great care in handling their contents which are often super personal. RFD isn't unique in this sense. In the zine librarian code of ethics, for example, the authors acknowledge "because of the highly personal content of zines, creators may object to having their material being publicly accessible" (Berthoud et al). In addition to being personal, zines are also counterculture productions. RFD and lots of other zines include descriptions of criminalized activity such as sex work, being queer, being unhoused, drug use, and acts of protest. Because of this, digitizing zines and making them searchable could potentially lead to the outing and/or arrest of their authors (who often used their real names and addresses).



This means that while the contents are online and freely available, you can only find a particular zine if you know what you're looking for. In some ways this is great. It means that cops can't just google someone's name and turn up a result on QZAP. But at the same time it limits other queers' abilities to find a particular piece of information unless it happens to be included in a metadata field or they know what they're looking for. The lack of OCR on QZAP is also a major accessibility issue because it means screen readers won't work on the page images.



This is a snapshot of QZAP's site featuring the Zine Naked R. of 2. All the meta data fields on the left are searchable. The PDF page image on the right is unsearchable.

History

RFD started in a country farmhouse in Grinnell Iowa in 1974, but its participants and readers came from across the US. The Iowa collective only ran the magazine for six issues. Thereafter, the Wolf Creek Oregon collective took over, then one in North Carolina, and so on and so forth until the current editor of upstate New York took on the editorial duties. Rather than try to take on an entire history of the magazine or fit it into one of those hegemonic heteropatriarchal models, Elizabeth and I decided we'd just focus on the history of the Iowa collective and the magazine's creation focusing mostly on the third issue because this one includes the fullest description of the editorial process. There are lot of gaps, but here's what we know.



It also means there's no clear way to know exactly how each issue was constructed unless the editors decided to record it. The Iowa collective only did this in the 3rd issue. So, from here on out we're going to focus on this 3rd issue's production history. This one's for spring 1975, and it's called Really Feeling Divine. Our editors are Dick, Don-Tevel, Ken, Kimmy, Larry, Olaf, Ollie, Rick, and Stewart from Iowa and Wisconsin. They gathered in an attic (location undisclosed) where they sorted through mailed in letters, photos, poems, illustrations, and stories. These could have been in color, in different type writer font or handwriting, etc. the editors typed of the written pieces, created some of their own, made the RFD map, and pasted it all together all while sipping coffee in an attic in an unknown locale.

Really Feeling Divine



Somewhere in Grinnell Iowa a collective of maybe 8 or so gay men lived in an old farmhouse. For about a year, they ran RFD. by "run the magazine," I don't mean much other than the Grinnell collective operated a PO box nearby the farmhouse to which people could mail contributions and handled the printing. The Grinnell collective had the Iowa City Women's Press print each issue on offset—more on that later. Aside from using ICWP as the printers, every issue has a totally unique production history, most of which are lost to time. the actual editors, tools, and design choices of each issue shifted. For example, for the second issue, the Iowa collective packed up their truck and traveled 2,000 miles to Eugene, OR, where they assembled the issue with gays from across the pacific northwest. Some issues feature a lot of potato stamp art—which was apparently the medium of choice of a frequent editor. Yet other issues feature no potato art at all (a real shame if you ask me). Most of the issues from this era are rendered using a typewriter with the same serif font (1, 2, 4, and 5) which may mean they were produced using the same typewriter, but I don't know for sure. Issues 3 and 6 have a sans serif font, but these fonts are distinct which probably means they came from two different typewriters. The point is, because a different group of people edited each issue, every issue is a unique embodied, queer production reflecting the bodies and minds of the different editors—some named, others not.

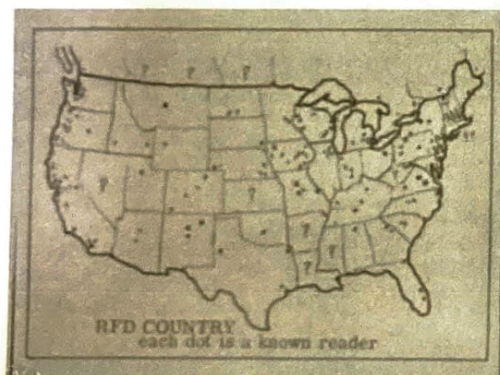




A lot of the gaps in the production history are caused by the fact that being gay was illegal in Iowa in 1974. This forced the editors to adopt a visibility politics distinct to rural queers.

RFD existed to forge a rural gay community which involved rural queers making themselves visible to each other. To that end, RFD was and still is a reader-driven magazine. Editors invited readers to submit letters, poems, art, short stories, recipes, essays, advertisements for their communes, really anything. The point, as one contributor put it, is to be "a place where people can come together and say: yeah, we do exist in some places" (Gavin and Dalton pg. 4 issue 2). People identified themselves by name (sometimes first and last) and even include their address so that other readers can write to them and drop by the next time they're on the road (people tended to be transient).

Contributors also sent in pictures of themselves, their partners, and their homes. In the Iowa issues, the editors created a map of the US on which they marked each location of known readers. The effect is a map that declares "we, a multitude of queer rural people, exist." Through making visible queer rural life, the editors managed to create a positive gay community in the face of what, on the day to day, was an isolating experience. Readers wrote in expressing their appreciation: "i really feel the mag. its a friend i can talk to" (John David pg. 2 is no. 2).



**RFD
MAP**

At the same time, the magazine's survival rested on its invisibility. For example, when the Grinnell collective founded RFD, same gender sex was still illegal in the state. In spite of this, issues contain images and illustrations of naked people, descriptions of gay sex work, and beautiful poems about friends and lovers. Suffice it to say, the production of RFD put the collective at constant risk of state sanctioned violence. For this reason, the Grinnell collective did not list their address, opting to use a PO box instead. Contributors who seem to be from the Grinnell collective never printed their last names (unlike those from other places).



RFD's rural queer visibility blurs the lines of public/private sphere that patriarchal conceptions of book history associate with the difference between published and unpublished work. In her 2017 article about fan fiction, Cait Coke suggests book historians "have accepted as a cultural norm hierarchies of value between print and digital that emphasize traditional patriarchal and public practices of reading and writing over private coterie practices, ones that have their roots in the history of women's reading and writing" (Coker). Coker historicizes digital to print fanfiction in the context of women's writing which has generally moved from private circles to sometimes more public ones. RFD's publication history is not dissimilar to that of a fanzine. Like digital or analog fanzines, RFD was created by shifting collectives of queers who gathered in private residences. While publicly available, RFD wasn't intended for public consumption. It was meant specifically for rural queers. And because of the illicit nature of the contents, the contributors balanced being both visible to each other and invisible to the state. Their visibility was both visible and invisible at the same time. In the sense that it resisted binarization, it was a distinctly queer kind of visibility.

This type of feminized publication dismantles this publish-for-public-consumption vs. write-for-private-viewing divide. It relied on concealing the identities of contributors to an outside public while rendering them visible to members. This provokes serious questions about the digitization of their often subversive contents.

IOWA CITY WOMEN'S PRESS

printed every issue the Iowa Collective took responsibility for including the 3rd. The Iowa City Women's Press printed a lot of countercultural (especially lesbian) publications over its 14 year existence. ICWP was a totally women (mostly lesbian) owned and operated press. It was founded in 1972 when the editors of the Iowa City Women's Liberation Front couldn't get an issue of their newspaper, *Ain't I A Woman?*, printed because it included instructions for DIY abortions (remember then abortion was illegal). So the women decided they'd print the newspaper themselves. From 1972-1988, ICWP printed countercultural publications that would never have been printed otherwise. RFD was one of these. It, like *Ain't I A Woman?*, contained illustrations and descriptions of criminalized embodiments and activities the state sought to eradicate.

ICWP began its operation in an old garage turned printing studio. They did their first publications on mimeograph, but "by June 1973 the icwp had an impressive amount of equipment, including a 10 yr. old 1250 [multilith printer], a 50 yr. old platen press, a NuArc platemaker (the old fashioned method), a Robertson vertical camera, a light table, a 30" papercutter, a hole punch (weight 1/2 ton), a folder, and a drafting table" (Beins and Enszer, 205-6). It seems likely that the printers over at ICWP use the collage-proofs from RFD to create a plate using the NuArc platemaker which they then fed into the 1250 multilith printer. What matters here is to note that the version of RFD that circulated looked totally different from that which exited the attic. This version is rendered almost entirely using different concentrations of black ink (with the exception of the cover which is in dark green). The thing that left the attic would have had some of its images in color. The pages are flat--there's nothing pasted on them--and they're all the same weight. And of course the substance of everything is the same--unlike the attic-borne piece which would have been made from typewriter ink, pen-ink, and collaged photos. That's not to say these page images lack depth. They don't. The different weights of the black make the two dimensional images leap from the page.

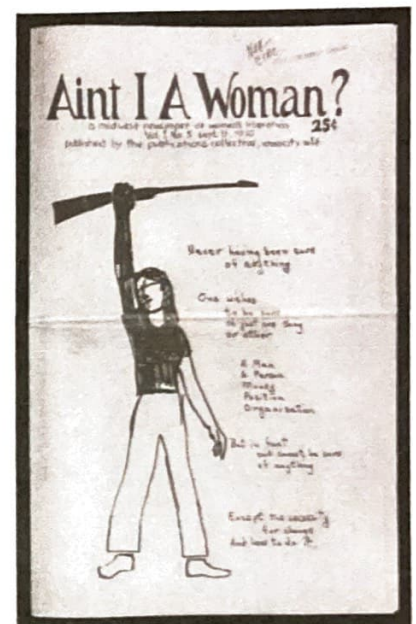
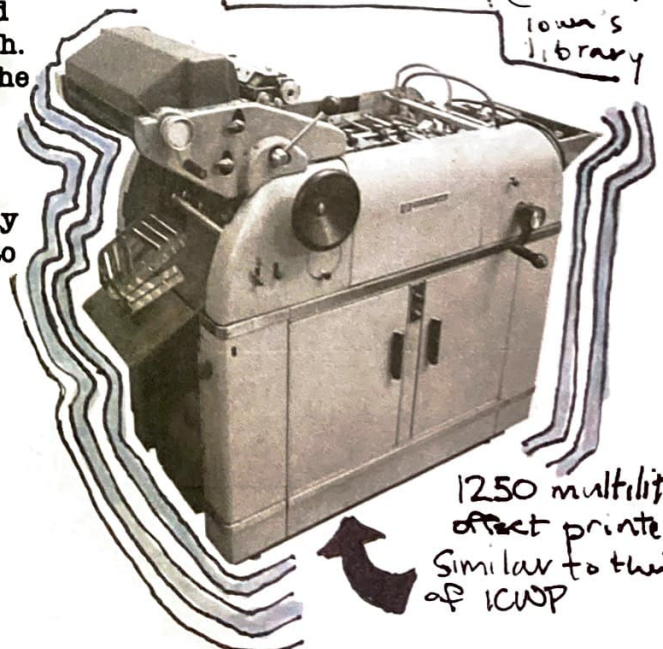


Image from 9/25/1970 7:35 AM of *Ain't I A Woman?* @ Univ of Iowa's library

Following the printing and binding, someone (possibly ICWP or the Grinnell collective) sent the magazines to readers across the country.



1250 multilith offset printer similar to that of ICWP

We conducted our OCR analysis with Tesseract version 5.1.0. We selected it because it is both open source and a very popular software for a variety of text types.

Tesseract is currently an open source project, but it was founded as a proprietary software by Hewlett-Packard. It was developed by Hewlett-Packard from the start of the project in 1985 until 1994. In 2005, HP made it an open source project, and then Google developed it from 2006 until 2018. Since then, development has continued as an open source project.

For this project, the open source style of Tesseract meant that the code was accessible to use, but that compiled versions of the software had to be found at different sources, and finding the newest version of Tesseract was not completely straightforward.

The Tesseract workflow starts by transforming the original image to black and white and then identifying the characters in the page. Version 5 has additional features in identifying different sections of the page, which improves its accuracy. RFD is written with many different page segments, which makes the section identification a very useful feature.

However, Tesseract still struggles to identify the parts of the page and order them correctly.

The open source nature of the Tesseract project, which distributes its history and materials across various groups and webpages, offered an interesting parallel to queer material like RFD. While Tesseract had its origins in the corporate world, it has since shifted or been pushed away from that. Where queer media was once pushed into the shadows, its symbols have been claimed by the corporate world, from Spotify playlists to Campbell's soup advertisements.

It's difficult to find much specific history about Tesseract. The details of its development are all documented; the git repo shows every modification made to the code. However, the overarching arch of the project is obfuscated in the details.

For example, they list Google as developing the project until 2018, but the dynamics about why that stopped are omitted.

Details about using Tesseract and how to retrain it with your own data are easy to find, but details about their training set are not available.

We downloaded the page images on which we ran Tesseract from RFD's issuu.com scan of issue 3: <https://issuu.com/rfdmag/docs/003> and ran the following shell script on them:

Executable File | 10 lines (7 sloc) | 127 Bytes

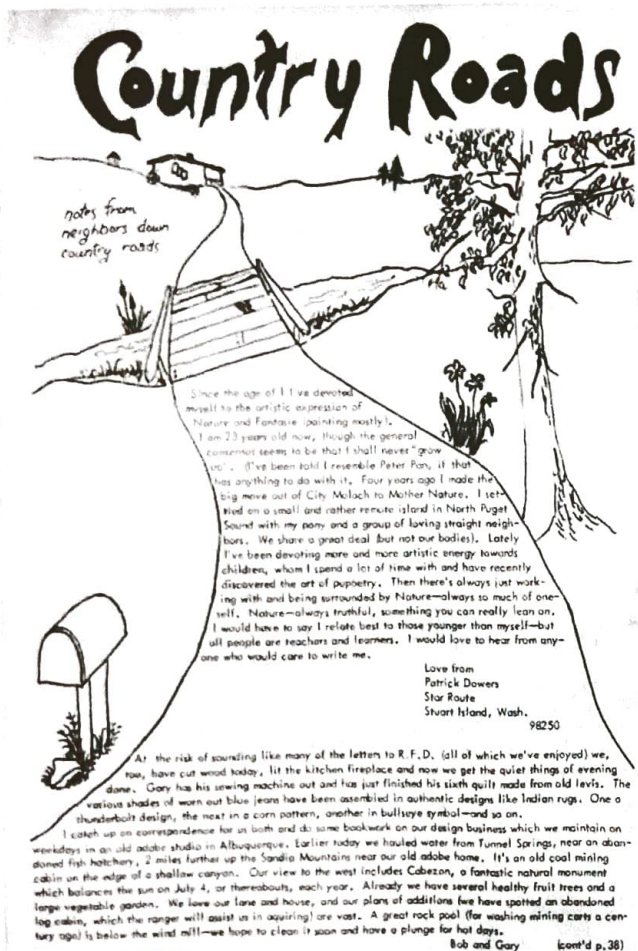
Raw Blame

```
1 #!/bin/bash
2
3 PRE=$1
4
5 for PAGE in $(ls -d $PRE/*.jpg)
6 do
7     OUT=$(echo $PAGE | sed "s/\\.jpg//g")
8     tesseract $PAGE $OUT
9 done
10
```

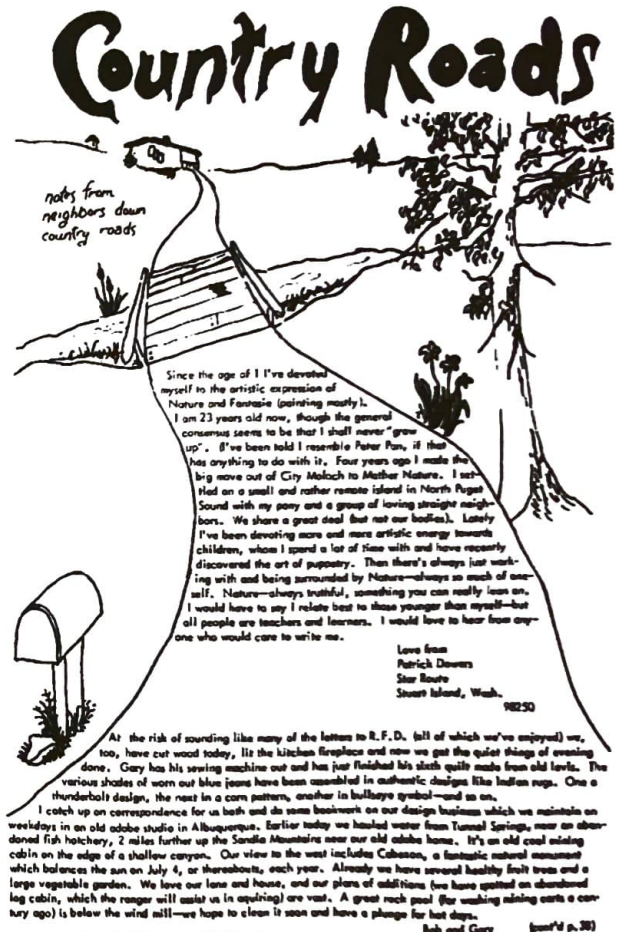
Tesseract

Output...

SCAN 2



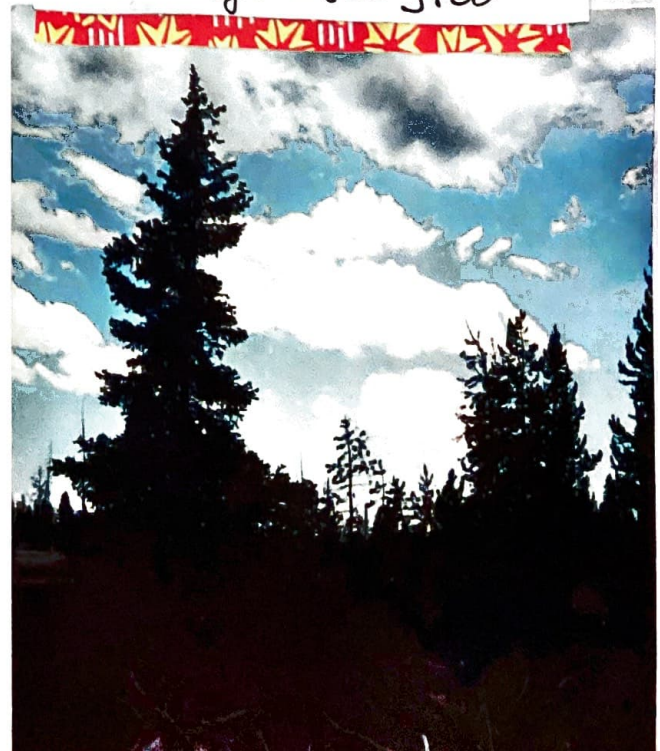
BINARIZED ↘



Binarization of the scan image.
Tesseract has to binarize the
image to match images to
characters.

The pages of RFD are an off-white, and the cover is green, but before Tesseract can execute its image to text process, it has to convert them to fully black and white. In some cases, this washes out the text. In others, it darkens the irregularities of the page colors. Regardless, it flattens the image to a binary grid.

Scan from www.ksnu.com RFD #3
Page 26. Tesseract ingests
this image as a JPEG



Since the age of 1 I've devoted myself to the artistic expression of Nature and Fantasie (painting mostly). I am 23 years old now, though the general consensus seems to be that I shall never "grow up". (I've been told I resemble Peter Pan, if that has anything to do with it. Four years ago I made the big move out of City Moloch to Mother Nature. I settled on a small and rather remote island in North Puget Sound with my pony and a group of loving straight neighbors. We share a great deal (but not our bodies), lately I've been devoting more and more artistic energy towards children, whom I spend a lot of time with and have recently discovered the art of puppetry. Then there's always just working with and being surrounded by Nature—always so much of oneself. Nature—always truthful, something you can really lean on. I would have to say I relate best to those younger than myself—but all people are teachers and learners. I would love to hear from anyone who would care to write me.

Love from
Patrick Dowers
Star Route
Stuart Island, Wash,
98250

For each of the letters it identifies on the page - whether a real or false letter - Tesseract takes its best guess for that letter. Often, it's correct. Whole paragraphs of text can have no more than a handful of errors. Other times, an image becomes gibberish or words happen in a ridiculous order. Even when it is "mostly" correct, Tesseract may repeat the same error in every sentence. Every time the word "I" occurred, Tesseract turned it into the pipe character. Although it can recognize many characters, and transcribe paragraphs, it can't identify the pronoun for the self.

'At the risk of sounding like many of the letters to R.F.D. (all of which we've enjoyed) we, too, have cut wood today, lit the kitchen fireplace and now we get the quiet things of evening done. Gary has his sewing machine out and has just finished his sixth quilt made from old levis. The Various shades of worn out blue jeans have been assembled in authentic designs like Indian rugs. One a thunderbolt design, the next in a corn pattern, another in bullseye symbol—and so on. I catch up on correspondence for us both and do some bookwork on our design business which we maintain on weekdays in an old adobe studio in Albuquerque. Earlier today we hauled water from Tunnel Springs, near an abandoned fish hatchery, 2 miles further up the Sandia Mountains near our old adobe home. It's an old coal mining cabin on the edge of a shallow canyon. Our view to the west includes Cabezón, a fantastic natural monument which balances the sun on July 4, or thereabouts, each year. Already we have several healthy fruit trees and a large vegetable garden. We love our lane and house, and our plans of additions (we have spotted an abandoned log cabin, which the ranger will assist us in acquiring) are vast. A great rock pool (for washing mining carts a century ago) is below the wind mill—we hope to clean it soon and have a plunge for hot days.

Bob and Gary (cont'd p.38)

If the images are represented at all by Tesseract, it is as a nonsense series of letters and punctuation. See the back cover for an example. RFD doesn't make sense without images. Even more so than the black and white version, the removal of word spacing and images removes the expressiveness from the original. It turns it from an interpersonal and relational creative work into a sequence of computerized characters.



27

Scan from pg 27



27

binarized scan from pg 27

ee

5 sts
: Seis

al

tesseract output plain text
of page 27

Tesseract, like any technology, has its affordances and limitations. It's visibility politics are totally built on the binarization of foreground and background, text and image, meaningfulness and meaninglessness.

Tesseract functions by running images through a binarization algorithm such that each pixel is rendered in black or white. This allows the model to separate text from background, things it deems meaningful from things it deems meaningless.

In the case of RFD, page images were already rendered in black and white. While the background of the pages for issue 3 are more of a cream than a black and white color, the binarization software did not dramatically change the page images much. Images remained as did most text. At this point, tesseract is still accounting for most of the page strictly because RFD was already "binarized" to an extent before we ran it through the program.

After tesseract binarizes page images, it runs the model on them to detect things it deems characters from other black marks that are non-textual. In this case, things that are text are preserved while the binarized images are cleaved away.

The results are better than expected. tesseract managed to detect most of the textual features of RFD and it did not render the non-textual/paratextual features as if they were text. We don't really know why this is the case, but we were pleasantly surprised. We collated this version of the issue and created a zine. Each page of the zine represents what tesseract outputted for the corresponding page of the printed issue.

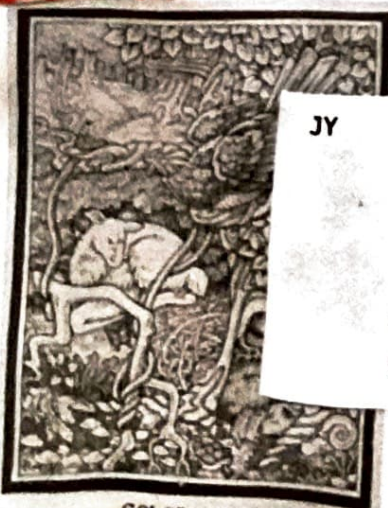
However, when I look at the printed out version of this pretty accurate transcription of RFD, I'm filled with an overwhelming sense of loss. In RFD that text/image binary that guides so much of computer vision doesn't exist. And frankly, I don't think it exists in other straighter publications either. That particular binarization/hierarchization of meaning leaves so much of how the contributors to RFD made each other visible to each other behind. Images of folks and their partners far by the wayside. Collage art of gays sitting in their geodesic domes aren't carried over. The participatory nature of the coloring page is replaced with the static-if hilarious-phrase:



Analysis of the Output

So much of the hand-crafted elements of the publication like potato stamps and cut out paper doll are lost. These pieces are just as important as the letters themselves in creating a living, breathing, record of rural gay existence. The bodily-ness of the zine is cleaved away by tesseract's hierarchization of text and image.

Tesseract makes the non-textual poetics of RFD imperceptible. In some ways this limit of the software keeps the invisible visibility of the magazine alive. Much like writers of RFD in the 70s made themselves visible to each other through entering the zine, but leaving out identifying information, our digitization might make the zine searchable without giving too much of its meaning away. Perhaps the major pitfalls of our rather flawed digitization are also its strength.



COLOR PAGE



COLOR PAGE



JY

C
OL
OR PA
G
E

We conclude this project with our dreams. Quantum dreams of a world beyond binarization. A world where multiplicities aren't smashed, contorted, and rendered into



black or white

Man or woman

Male or female

True or False

Gay or straight

0 or 1