

ORIGINAL

IN THE SUPREME COURT OF OHIO

STATE, ex rel. DATA TRACE
INFORMATION SERVICES, LLC, et al.,

Case No. 10-2029

Relators,

vs.

ORIGINAL
ACTION IN MANDAMUS

RECORDER OF CUYAHOGA
COUNTY, OHIO,

Respondent.

BRIEF OF AMICUS CURIAE OHIO NEWSPAPER ASSOCIATION
IN SUPPORT OF RELATORS

David Marburger* (0025747)

**Counsel of Record*

John Blanton (0086836)

Baker Hostetler LLP

3200 PNC Center

1900 East 9th Street

Cleveland, OH 44114

Phone: (216) 621-0200

Fax: (216) 696-0740

dmarburger@bakerlaw.com

jblanton@bakerlaw.com

Counsel for Relators

David T. Movius* (0070132)

**Counsel of Record*

Matthew Cavanagh (0079522)

MCDONALD HOPKINS LLC

600 Superior Avenue, East, Suite 2100

Cleveland, OH 44114

Phone: (216) 348-5400

Fax: (216) 348-5474

dmovius@mcdonaldhopkins.com

mcavanagh@mcdonaldhopkins.com

Counsel for Respondent

John C. Greiner (0005551)

GRAYDON HEAD & RITCHEY LLP

1900 Fifth Third Center

511 Walnut Street

Cincinnati, OH 45202-3157

Phone: (513) 629-2734

Fax: (513) 651-3836

jgreiner@graydon.com

Counsel for Amicus Curiae

Ohio Newspaper Association

FILED
JUN 15 2011
CLERK OF COURT
SUPREME COURT OF OHIO

TABLE OF CONTENTS

	<u>Page</u>
TABLE OF AUTHORITIES	ii
I. STATEMENT OF THE CASE	1
II. ARGUMENT	1
A. The Value of Database Journalism	1
B. Ohio Courts Have Consistently Recognized Databases As Public Records	3
C. Ohio Attorney General Opinions Have Consistently Limited the Scope of R.C. 317.32(I)	5
III. CONCLUSION	7
CERTIFICATE OF SERVICE	8
SUPPLEMENT (1-33)	
Bowen, Ezra, <i>Press: New Paths to Buried Treasure</i> , Time (July 7, 1986), http://www.time.com/time/printout/0,8816,961680,00.html	1
Cohen, Sarah, <i>Computational Journalism: A Call to Arms to Database Researchers</i> , Conference on Innovative Data Systems Research (2011), http://www.cidrdb.org/cidr2011/program.html	3
Mikelbank, Brian, <i>Beyond the Foreclosure Crisis Housing Strategy for Cleveland's Future</i> , Center for Housing Research and Policy (September 2008), www.urban.csuohio.edu/housing	7
Affidavit of Jayson Gerbec	29

TABLE OF AUTHORITIES

CASES

<i>State ex rel. Athens Cty. Property Owners Assn. v. City of Athens</i> (1992), 85 Ohio App.3d 129, 619 N.E.2d 437	4, 5
<i>State ex rel. Cincinnati Post v. Schweikert</i> (1988), 38 Ohio St.3d 170, 527 N.E.2d 1230	3, 4
<i>State ex rel. Margolius v. City of Cleveland</i> (1992), 62 Ohio St.3d 456, 584 N.E.2d 665	4, 5
<i>State ex rel. Strothers v. Wertheim</i> (1997), 80 Ohio St. 3d 155, 684 N.E. 2d 1239	1

STATUTES

R.C. 149.43	1, 7
R.C. 149.43(B).....	6
R.C. 317.32(I).....	1, 5, 6

OTHER

1994 Ohio Atty.Gen.Ops. No. 94-006	6
2004 Ohio Atty.Gen.Ops. No. 2004-001	6
2004 Ohio Atty.Gen.Ops. No. 2004-033	6, 7
Bowen, Ezra, <i>Press: New Paths to Buried Treasure</i> , Time (July 7, 1986), http://www.time.com/time/printout/0,8816,961680,00.html	2
Cohen, Sarah, <i>Computational Journalism: A Call to Arms to Database Researchers</i> , Conference on Innovative Data Systems Research (2011), http://www.cidrdb.org/cidr2011/program.html	3
Mikelbank, Brian, <i>Beyond the Foreclosure Crisis Housing Strategy for Cleveland's Future</i> , Center for Housing Research and Policy (September 2008), www.urban.csuohio.edu/housing	2

I. STATEMENT OF CASE

This case presents a straight forward question. Should Ohio's Public Records Act¹ ("PRA"), which is to be interpreted broadly to promote governmental transparency², give way to a statute that imposes a prohibitive charge for photocopies,³ when that statute has no application to the particular request? In more direct terms, can a county recorder invoke a statute calling for a \$2.00 per page charge for a photocopy when the requester asks merely for an electronic copy of data stored on a compact disk ("CD")?

The answer to both questions is "no." And while this case, in its most narrow terms, concerns entities involved in the real estate title business, its broader application has a direct and enduring impact on journalism. The Ohio Newspaper Association provides this Amicus brief to ensure that the Court protects the vital interests of journalists and ultimately the public.

II. ARGUMENT

A. The Value of Database Journalism.

The very term "investigative journalism," suggests what it involves. Throughout our country's history, journalists, acting initially on "tips," have painstakingly gathered information from a number of diverse sources and exposed misfeasance, malfeasance and corruption.

One source of information that has been invaluable in the investigative process is the database. In its simplest terms a "database" is a collection of a large amount of information in a single, searchable source. There are two reasons why databases are such valuable tools for investigative journalists. First, the availability of voluminous amounts of data in a readily

¹ R.C. 149.43 et seq.

² "Time and time again we have held that 149.43 must be construed liberally in favor of broad access and any doubt should be resolved in favor of disclosure of public records." See *State ex rel. Strothers v. Wertheim* (1997), 80 Ohio St.3d 155, 156, 684 N.E.2d 1239.

³ R.C. 317.32(I).

searchable format allows a journalist to spot patterns that expose fraud, waste and other wrongdoing. Thus for example:

“In Providence, the Journal-Bulletin used an IBM 4381 to analyze 30,000 low-interest mortgages issued by the Rhode Island Housing and Mortgage Finance Corp. By matching mortgage issue dates with the bond issue that financed them, the machine helped expose a ‘secret fund’ that apparently was used to give out RIHMFC loans to politically connected people like the daughter of former Governor J. Joseph Garrahy.... Following the revelations in the Journal-Bulletin, the Fleet National Bank and 24 individuals were indicted.”⁴

“In 1979, for instance, the Miami Herald scanned with a computer all 2 million of Dade County’s property-tax assessments to dig out inequities. In 1984 Long Island’s (N.Y.) Newsday parsed every state-awarded highway contract in the area and all major county sewer contracts over eleven years to discover that five favored firms collected 86% of the boodle.”⁵

More recently, researchers in Cuyahoga County, Ohio reviewed 47,000 mortgages issued from 2003 through the third quarter of 2007 to expose the high incidence of negative equity in certain Cleveland wards.⁶ The report then used this information to make recommendations for addressing the foreclosure crisis.⁷

Second, the ability to quickly search a database containing thousands, if not millions of information bits is absolutely essential in today’s climate of dwindling journalistic resources. According to some reports, full-time newsroom employment has fallen by 25% over the past ten

⁴ Bowen, Ezra, *Press: New Paths to Buried Treasure*, Time (July 7, 1986), <http://www.time.com/time/printout/0,8816,961680,00.html>.

⁵ *Id.*

⁶ Mikelbank, Brian, *Beyond the Foreclosure Crisis Housing Strategy for Cleveland's Future*, Center for Housing Research and Policy (September 2008), www.urban.csuohio.edu/housing.

⁷ *Id.*

years.⁸ Any tool that allows journalists to do more with less is a necessity, not a luxury, in this new world. Given the commitment to transparency reflected in the PRA, courts must encourage the availability of databases, rather than discourage it.

The Cuyahoga County Recorder's insistence on charging \$2.00 per page for images stored on a CD very much discourages the availability of a valuable database. The Affidavit of Jayson Gerbec⁹ illustrates the point. After initially being misinformed about the Recorder's practice of maintaining records on CDs, Mr. Gerbec received an e-mail from Larry Patterson at the Recorder's office informing him that the charge for copying records onto 4 CDs (which Mr. Gerbec supplied) would total \$35,568.¹⁰

A \$35,000 copy charge renders the database off limits to the press and the public. Given that the actual cost to make the copy is close to zero, there is no way to justify the recorder's practice.

B. Ohio Courts Have Consistently Recognized Databases As Public Records.

As far back as 1988, the Ohio Supreme Court recognized that "compilations" of public records are themselves public records, subject to production pursuant to the PRA. In *State ex rel. Cincinnati Post v. Schweikert*,¹¹ this court recognized that compilations of statistics about individuals in jail and individuals about to be sentenced were records subject to the PRA.¹² The fact that the compilations represented the work product of the court administrator and were compiled from publicly available information did not alter the analysis.¹³ This Court put it best when it said: "The law does not require members of the public to exhaust their energy and

⁸ Cohen, Sarah, *Computational Journalism: A Call to Arms to Database Researchers*, Conference on Innovative Data Systems Research (2011), <http://www.cidrdb.org/cidr2011/program.html>.

⁹ Affidavit of Jayson Gerbec.

¹⁰ *Id.* at ¶¶ 12-16.

¹¹ *State ex rel. Cincinnati Post v. Schweikert* (1988), 38 Ohio St.3d 170, 527 N.E.2d 1230.

¹² *Id.* at 174.

¹³ *Id.* at 173.

ingenuity to gather information which is already compiled and organized in a document created by public officials at public expense.”¹⁴

This court expanded on the *Schweikert* holding in *State ex rel. Margolius v. City of Cleveland*.¹⁵ In *Margolius*, the relator sought copies of computer tapes detailing activity of the Cleveland Police Department from 1980 to the date of the request.¹⁶ The city insisted on providing the data on paper printouts.¹⁷ The relator asked the court to order the city to produce a copy of the computer tape, noting that the printouts would fill 13 boxes, and be unduly difficult to transfer to computer files for analysis.¹⁸

Once again, this court put substance over form and ordered the city to produce a copy of the tape.¹⁹ It noted:

“As this example illustrates, the manner in which the records are organized can add to the value of the information contained within those records. When such value is added, a new set of enhanced public records is created that must be disclosed to the public.... Here, the added value is not only the organization of the data, but also the compression of the data into a form that allows greater ease of public access.”²⁰

The Fourth District Court of Appeals applied the Supreme Court precedent in *State ex rel. Athens Cty. Property Owners Assn. v. City of Athens*.²¹ There, the relator sought certain records concerning rental property, and asked the City of Athens to provide the information on

¹⁴ *Id.*

¹⁵ *State ex. rel. Margolius v. City of Cleveland* (1992), 62 Ohio St.3d 456, 584 N.E.2d 665.

¹⁶ *Id.* at 458.

¹⁷ *Id.*

¹⁸ *Id.*

¹⁹ *Id.* at 460.

²⁰ *Id.*

²¹ *State ex. rel. Athens Cty. Property Owners Assn. v. City of Athens* (1992), 85 Ohio App.3d 129, 619 N.E.2d 437.

computer diskettes.²² While acknowledging that the records were subject to the PRA, *Athens* nonetheless insisted on producing hard copies.²³ The court rejected *Athens*' position, noting:

“The record shows that the records are normally stored on an electronic medium, that those records are compiled using taxpayer dollars, on equipment purchased with taxpayer dollars. The record also shows that the requested information consists of over six hundred records and that the ACPOA [relator] would have to go to needless expense to replicate these records from hard copy. *Athens*' assignment of error is not well taken and is overruled.”²⁴

The Respondent's position conflicts with all of the aforementioned cases. The Respondent is essentially saying that, despite the fact that the records are electronically compiled, it is entitled to disregard that fact and act as though it is providing photocopies of paper records. In that respect, its position is weaker than the respondents in the *Margolius* and *Athens* cases. In those cases, the respondents insisted that relators accept the records in hard copy. In this case, the Respondent is insisting that relators **pretend** they are getting hard copies. If this court rejected the respondents' arguments in *Margolius* and *Athens*, it **must** reject Respondent's argument here, especially since that argument requires the Court to engage in make believe.

C. **Ohio Attorney General Opinions Have Consistently Limited the Scope of R.C. 317.32(I).**

Respondent seeks to apply R.C. 317.32(I) in an incredibly expansive fashion. It would impose a statutory per page “photocopy” charge to the process of electronically copying a CD. The Ohio Attorney General has consistently issued opinions, however, which demonstrate that there is no basis for such a broad interpretation.

²² *Id.* at 130.

²³ *Id.*

²⁴ *Id.* at 132.

Thus in 1994, the Ohio Attorney General responded²⁵ to the following question: “If a microfiche has a hundred documents or a roll of film contains five hundred documents, can a member of the public pay for the cost to reproduce such fiche or roll rather than pay the \$1.00 per page charge under [R.C. 317.32(I)]?”²⁶

The Attorney General, applying the *Margolius* and *Schweikert* cases, opined: “If a person requests copies of public records stored by the county recorder on microfiche or film, R.C. 149.43(B) requires the county recorder to make available in the same medium a copy of the portions of the microfiche or film containing those public records, if the person requesting such copy has presented a legitimate reason why a paper copy of the records would not be sufficient or practicable, and if the person assumes the expense of making a copy in that medium, in lieu of the photocopying fee prescribed by R.C. 317.31(I).”²⁷

This case presents the very same question, albeit regarding updated technology. And there is no reason for this Court to arrive at a different conclusion.

In two more recent opinions, the Attorney General declined to adopt the expansive interpretation the Respondent now offers. Thus, in Opinion 2004-011, the Attorney General reasoned that where a member of the public used a digital camera to make copies of recorder’s records the recorder could not charge the R.C. 317.32(I) per page fee.²⁸ And in Opinion 2004-033, the Attorney General opined that where a county recorder maintained a photocopy machine in its office, but no staff assisted the public in making the copies, the recorder could not charge

²⁵ 1994 Ohio Atty.Gen.Ops. No. 94-006.

²⁶ *Id.*

²⁷ *Id.*

²⁸ 2004 Ohio Atty.Gen.Ops. No. 2004-001.

the \$2.00 per page fee. In that case, the recorder could charge only for its "cost" as provided in R.C. 149.43.²⁹

These Opinions establish, without question, that R.C. 317.21(I) means only what it says and no more. Where the recorder or its staff actually makes a photocopy of a record, it may charge the \$2.00 per page fee. In **any** other scenario, R.C. 317.21(I) does **not** apply and the recorder may only charge its actual cost as set forth in R.C. 149.43. This is consistent with common sense, the English language, and the law that compels a broad interpretation of R.C. 149.43.

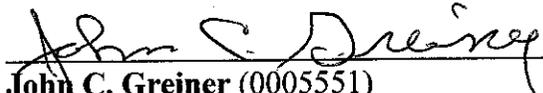
This case is not terribly complicated. The Cuyahoga County Recorder is **not** making a photocopy of any record. And it cannot, therefore, charge a fee that would make a valuable database inaccessible to the public.

IV. CONCLUSION

This court should take the opportunity presented by this case to advance the public's right to a transparent government. Where available technology enhances that right, this court should not apply the Revised Code in an antiquated manner that frustrates the advantages provided by technological advances.

²⁹ 2004 Ohio Atty.Gen.Ops. No. 2004-033.

Respectfully submitted,



John C. Greiner (0005551)
GRAYDON HEAD & RITCHEY LLP
1900 Fifth Third Center
511 Walnut Street
Cincinnati, OH 45202-3157
Phone: (513) 629-2734
Fax: (513) 651-3836
jgreiner@graydon.com

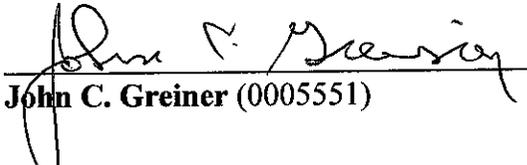
*Counsel for Amicus Curiae
Ohio Newspaper Association*

CERTIFICATE OF SERVICE

The undersigned hereby certifies that a true and accurate copy of the foregoing was served by regular U.S. Mail, postage prepaid, this 15th day of June, 2011, upon the following:

David T. Movius, Esq.
Matthew Cavanagh, Esq.
MCDONALD HOPKINS LLC
600 Superior Avenue, East
Suite 2100
Cleveland, OH 44114

David Marburger, Esq.
John Blanton, Esq.
BAKER HOSTETLER LLP
3200 PNC Center
1900 East 9th Street
Cleveland, OH 44114



John C. Greiner (0005551)




SWAP YOUR RIDE
SALES EVENT

2012 Focus. UP TO **40** hwy mpg*

Roll over to expand and see their stories

[Back to Article](#)

[Click to Print](#)



Monday, Jul. 07, 1986

Press: New Paths to Buried Treasure

By Ezra Bowen.; Thomas McCarroll/New York and Jonathan Wells/Providence

Traditionally, investigative reporters have relied on shoe leather and patience to ferret out the proverbial smoking gun. However, in the computer age, newsmen are enlisting the machine with dramatic results:

In Providence, the Journal-Bulletin used an IBM 4381 to analyze 30,000 low-interest mortgages issued by the Rhode Island Housing and Mortgage Finance Corp. By matching mortgage issue dates with the bond issue that financed them, the machine helped expose a "secret fund" that apparently was used to give out RIHMFC loans to politically connected people like the daughter of former Governor J. Joseph Garrahy (who has since returned the loan). Following the revelations in the Journal-Bulletin, the Fleet National Bank and 24 individuals were indicted.

At the Milwaukee Journal, Reporter Nina Bernstein heard that undefended indigents were being jailed for months because they could not pay \$100-to-\$300 fines for offenses like jaywalking. She went after the story, helped by clerks who fed records of 899 inmates through a computer. Says Bernstein: "I interviewed the judges last and presented them with the evidence, and they were stunned." The courts freed hundreds of inmates, threw out 20,000 orders for jail commitment and told the county to provide attorneys for poor defendants. The computer's statistics, says Bernstein, "made our case airtight." She adds that without the machine, her work would have required up to a year of research rather than the six weeks it took.

In 1985 Reporter Richard Mauer of the Anchorage Daily News joined with Computer Wizard and Freelance Reporter Larry Makinson to trace campaign contributions given to local officials. They uncovered a nest of questionable schemes, including one to funnel 20 seemingly independent \$1,000 contributions to a single state senator in one day. "Without the computer," says Makinson, "this information would have remained buried like a treasure chest at the bottom of the sea."

This week Bernstein and the Milwaukee Journal as well as Makinson and Mauer of the Anchorage Daily News are being recognized for their high-tech digging! with awards from an association called Investigative Reporters & Editors, with headquarters at the University of Missouri School of Journalism.

Says I.R.E. Executive Director Steve Weinberg: "The computer is revolutionizing investigative reporting. There's just no way you could do some of those calculations by hand."

The acknowledged guru of the computer movement is Philip Meyer, 55, now professor of journalism at the University of North Carolina at Chapel Hill. Meyer first used a computer as an investigative tool when he was a reporter for the Detroit Free Press, analyzing the demographics of blacks in Detroit's 1967 riots. He had previously worked on a computer while on a Nieman Fellowship at Harvard. Says Meyer: "Harvard had an IBM 7090, and I learned to apply it to social science." Meyer's findings on the riots helped the Free Press win a Pulitzer. It also inspired him to write *Precision Journalism*, a computer reporters' bible that came out in 1973. Among the first reporters to turn to the machine were the Philadelphia Inquirer's Donald Barlett and James Steele. They used an IBM for a 1973 series that won two national awards for revealing disparate court sentencing of violent offenders. "If we did that story all by hand," says Steele, "we'd still be working on it." But, he cautions, "the computer does not take the place of traditional reporting, analysis or pavement pounding. It's just another tool."

The prime uses: archival scanning that once required exhaustive card- catalog searches and high-speed analysis of myriad numbers until the machine kicks out revelatory patterns. In 1979, for instance, the Miami Herald scanned with a computer all 2 million of Dade County's property-tax assessments to dig out inequities. In 1984 Long Island's (N.Y.) Newsday parsed every state- awarded highway contract in the area and all major county sewer contracts over eleven years to discover that five favored firms collected 86% of the boodle.

The plethora of public records that local governments now store on magnetic tape for their own computers greatly widens the field for inquiring reporters. The Providence Journal-Bulletin has compiled a library of computer tapes that includes the record of every criminal defendant who has appeared in Rhode Island Superior Court in the past nine years, as well as the state's entire fiscal records for 2 1/2 years. With this data base, the paper has uncovered coercive tactics used by some canvassers in the state's mail-in electoral ballots and has revealed that a total of 6,033 arsons in 1982 resulted in only 19 jail sentences. "The power of this thing is unbelievable," says Journal- Bulletin Reporter Elliot Jaspin. "Newspapers are either going to start doing what we do, or they're going to be bypassed and out of date."

 [Click to Print](#)

Find this article at:

<http://www.time.com/time/magazine/article/0,9171,961680,00.html>

Copyright © 2011 Time inc. All rights reserved. Reproduction in whole or in part without permission is prohibited.
[Privacy Policy](#) | [Add TIME Headlines to your Site](#) | [Contact Us](#) | [Customer Service](#)

Computational Journalism: A Call to Arms to Database Researchers

Sarah Cohen
School of Public Policy
Duke Univ.

sarah.cohen@duke.edu

Chengkai Li
Dept. of Comp. Sci. & Eng.
Univ. of Texas at Arlington

cli@uta.edu

Jun Yang
Dept. of Comp. Sci.
Duke Univ.

junyang@cs.duke.edu

Cong Yu
Google Inc.

congy@umich.edu

1. INTRODUCTION

The digital age has brought sweeping changes to the news media. While online consumption of news is on the rise, fewer people today read newspapers. Newspaper advertising revenues fell by a total of 23% in 2007 and 2008, and tumbled 26% more in 2009 [1]. This continuing decline of the traditional news media affects not only how news are disseminated and consumed, but also how much and what types of news are produced, which have profound impact on the well-being of our society. In the past, we have come to rely heavily upon the traditional news organizations for their investigative reporting to hold governments, corporations, and powerful individuals accountable to our society. The decline of traditional media has led to dwindling support for this style of journalism, which is considered as cost-intensive and having little revenue-generating potential.

Today, there are fewer reporters gathering original material than in a generation. By some reports, full-time newsroom employment has fallen by one-quarter over the past ten years [2]. Bloggers, citizen journalists, and some non-profit news agencies have made up for only a small part of this loss. The growth in the online news organizations has been mostly in the role of aggregators, who read other blogs and news reports, and select, aggregate, edit and comment on their findings. There is a real danger that the proud tradition of original, in-depth investigative reporting will fade away with the ailing traditional news media.

Luckily, a second trend is on our side: the continuing advances in computing. We are connecting people together on unprecedented scales. Data collection, management, and analysis have become ever more efficient, scalable, and sophisticated. The amount of data available to the public in a digital form has surged. Problems of increasing size and complexity are being tackled by computation. Could computing technology—which has played no small part in the decline of the traditional news media—turn out to be a savior of journalism's watchdog tradition?

In the summer of 2009, a group (including two authors of this paper) of journalists, civic hackers, and researchers in social science and computer science gathered for a workshop at Stanford on the nascent field of *computational journalism*, and discussed how computation can help lower cost, increase effectiveness, and encourage participation for investigative journalism. In this paper,

This article is published under a Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0/>), which permits distribution and reproduction in any medium as well allowing derivative works, provided that you attribute the original work to the author(s) and CIDR 2011.

5th Biennial Conference on Innovative Data Systems Research (CIDR '11)
January 9-12, 2011, Asilomar, California, USA.

we present a more focused perspective from database researchers by outlining a vision for a system to support mass collaboration of investigative journalists and concerned citizens. We discuss several features of the system as a sample of interesting database research challenges. We argue that computational journalism is a rich field worthy of attention from the database community, from both intellectual and social perspectives.

2. A CLOUD FOR THE CROWD

News organizations today have little time and resources for investigative pieces. It is economically infeasible for most news organizations to provide their own support for investigative journalism at a sufficient level. To help, we envision a system based on a *cloud for the crowd*, which combines computational resources as well as human expertise to support more efficient and effective investigative journalism.

The “cloud” part of our vision is easy to understand. Treating computing as a utility, the emerging paradigm of cloud computing enables users to “rent” infrastructure and services as needed and only pay for actual usage. Thus, participating news units can share system setup and maintenance costs, and each reporter can access a much bigger pool of resources than otherwise possible. Popular tools, such as those supporting the Map/Reduce model, have made scalable data processing easy in the cloud. These tools are a perfect fit for many computational journalism tasks that are inherently data-parallel, such as converting audio or scanned documents to text, natural language processing, extracting entities and relationships, etc. Finally, sharing of infrastructure and services encourages sharing of data, results, and computational tools, thereby facilitating collaboration.

Cloud for computational journalism is becoming a reality. A pioneering example is DocumentCloud.org, started by a group of journalists at ProPublica and the *New York Times* in 2008. It hosts original and user-annotated documents as well as tools for processing and publishing them. Conveniently, it uses a Ruby-based Map/Reduce implementation to perform document OCR on Amazon E2. Going beyond DocumentCloud's document-centricity, the system we envision would also help manage, integrate, and analyze *structured data*, which may be either extracted from text or published by a growing number of public or government sources. With structured data, we can draw upon a wealth of proven ideas and techniques from databases, ranging from declarative languages, continuous querying, to parallel query processing. Implementing them in the cloud setting raises new challenges, and is a direction actively pursued by the database community. Computational journalism may well be a “killer app” for this line of research.

We would like to emphasize the “crowd” part of our vision more, however. While the “cloud” part copes with the need for

computational resources, investigative journalism will never be fully automated by computation. How can we leverage the “crowd” to cope with the need for human expertise? As a start, DocumentCloud allows sharing of human annotations on documents and facilitates collaborative development of tools. With declarative information extraction and declarative querying, we could further share and reuse results (final or intermediate) of data processing tasks initiated by different users. From the perspective of the system, such collaboration occurs opportunistically. Ideally, we want to *actively* direct the efforts of the crowd. For example, in 2009, *The Guardian* of London put up close to half a million pages of expense documents filed by British MPs on the Web, and asked viewers to help identify suspicious items. Within 80 hours, 170,000 documents were examined, making this crowdsourcing effort a spectacular success [3]. As an example from the database community, the *Cimple* project on community information management [4] relies on user feedback to verify and improve accuracy of information extracted from Web pages. Many existing approaches assign jobs to the crowd in simple ways (e.g., pick any document to check). They will certainly get the job done in the long run, but if we have in mind an idea for a story with a tight deadline—which is often the case for journalists—assigning many jobs whose outcomes bear little relevance to our goal will dilute the crowd’s efforts.

What we envision instead is a system that intelligently plans (or helps to plan) the crowd’s efforts in a *goal-driven* fashion. Given a computational task (e.g., a query), the system would generate a tentative result based on the data it currently has. At this point, the result can be quite uncertain, because the data contain all sorts of uncertainty, ranging from imperfect information extraction to errors in publicly released datasets. Suppose the result looks newsworthy, and a journalist is allocated a limited amount of time and resources to investigate this lead. To leverage the power of the crowd, our system would come up with mini-tasks to be crowdsourced. These mini-tasks can be as simple as checking an entity-relationship extracted from a document, or as complex as reconciling entries from two different public databases. Different lists of mini-tasks would be presented to different users according to their expertise and preference. Mini-tasks whose completion contributes most to reducing the overall result uncertainty will be listed as having a higher priority. As results of mini-tasks become available, the system would reevaluate the overall result, and adjust and reprioritize the remaining mini-tasks accordingly. This idea can be seen as a generalization of the pay-as-you-go approach to data integration in *dataspaces* [5], which considered mini-tasks that establish correspondences between entities from different data sources.

To better illustrate the new challenges and opportunities involved, put yourself in the shoes of a journalist who just noticed a huge number of blog posts about high crime rates around the Los Angeles City Hall. First, are there really that many posts about high crime rates in this area, or did the automated extraction procedure pick up something bogus? Second, does having a large number of blog posts necessarily increase the credibility of the claim, or did most of these posts simply copy from others? Knowing that the number of *original* sources for a story is almost always very low, a seasoned journalist will likely start with a few popular posts, verify that they indeed talk about high crime rates around Los Angeles City Hall, and then trace these posts back to find the original sources. When planning for crowdsourcing, our system should try to mimic the thought process of seasoned

journalists. In particular, it would be suboptimal to assign mini-tasks for verifying extraction results from a lot of posts, because the number of posts making a claim is a poor indicator for the accuracy of the claim anyway, and checking just a few may boost our confidence in the extraction procedure enough. It is also suboptimal to ask the crowd to trace the sources of many posts, because a handful of them may lead us to the few original sources.

In this case, it came down to a couple of sources: a Los Angeles Police Department site for tracking crimes near specific addresses, and EveryBlock.com, which publishes large bodies of public-domain data (such as crimes and accidents) by location and neighborhood. Further investigation reveals that EveryBlock.com republished data from LAPD, so our task reduces to that of verifying the claim in the LAPD database (a topic that we shall return to in Section 3). Interestingly, according to the geocoded map locations of crimes in the database, the numbers check out: the crime rate at 90012, ZIP code for the Los Angeles City Hall, indeed ranked consistently as the highest in the city. But a true investigative journalist does not stop here; in fact, there is where the fun begins. It would be nice for our system to help journalists quickly eliminate other possibilities and zoom in on the fun part.

As it turned out, there was a glitch in the geocoding software used by the LAPD site to automatically convert street addresses to map locations. Whenever the conversion failed, the software used the default map location for Los Angeles, right near the City Hall, hence resulting in a disproportionately high crime rate. Arriving at this conclusion does require considerable skill and insight, but our system can help by providing easy access to the full dataset (with street addresses included), and by crowdsourcing the mini-tasks of checking crime records that have been geocoded to the default location (if no alternative geocoding software is available).

Much of what we described in this example took place in real life, and was the subject of a 2009 story in the *Los Angeles Times* [6]. We do not know how many bloggers picked up the false information, but the fact that the *Los Angeles Times* published this piece about the software glitch instead of a column on crimes near the City Hall is both comforting and instructive. As the Internet has made it trivial to publish (and republish) information—especially with the proliferation of social networking—there is a real danger of misinformation going viral. It is important for computational journalism to help preserve journalistic principles and to facilitate fact-checking (more on these in Section 3).

Many research challenges lie ahead of us in supporting intelligent planning of crowdsourcing for investigative journalism. Before we can hope to replicate or improve the cost-benefit analysis implicitly carried out in the mind of a seasoned journalist, we first need frameworks for representing prior knowledge and uncertainty in data (raw and derived) and reasoning with them. There has been a lot of work on probabilistic databases [7], and it would be great to put the techniques to a serious test. For example, how do we represent a large directed acyclic graph of uncertain dependencies among original sources and derived stories? How do we capture the belief that the number of original sources is small? We believe studying our application will necessitate advances in data uncertainty research.

Given a declarative specification of what we seek from data, we also need methods to determine what underlying data matter most to the result. Akin to sensitivity analysis, these methods are crucial in prioritizing mini-tasks. Work on lineage [8] took an

important initial step towards this direction, but we are interested in not only *whether* something contributes to the result, but also *how much* it would affect the result when it turns out to be something else. Work on dataspace [5] laid out an interesting direction based on the concept of the *value of perfect information*, but with general mini-tasks and more complex workflows involving extraction, cleansing, and querying, the problem becomes more challenging.

In addition to the benefit of a mini-task, we will also need to quantify its cost. The idea of exploring the cost-benefit tradeoff has been investigated in the context of *acquisitional query processing* in sensor networks [9]. The human dimension of the crowd creates many new problems. Interests, expertise, and availability vary greatly across users. Some mini-tasks may never be picked up. Sometimes users do a poor job. Therefore, it is difficult to predict a mini-task's cost and result quality (which affects its *actual* benefit). Also challenging are the problems of adjusting crowdsourcing plans dynamically based on feedback, coordinating crowdsourcing among concurrent investigations, and allocating incentives using, say, the Amazon Mechanical Turk. Efforts such as American Public Media's Public Insight Network have taken a qualitative approach toward building and using participant profiles. It would be interesting to see whether a more quantitative approach can be made to work for a crowd.

To recap, we envision a system for computational journalism based on "a cloud for crowd," which hosts tools and data (raw and derived, unstructured and structured), runs computational tasks, and intelligently plans and manages crowdsourcing. It combines resources and efforts to tackle large tasks, and it seeks to leverage and augment human expertise. Within the system, there are endless possibilities for innovative applications of computing to journalism. In the next section, we describe a few specific ideas with clear database research challenges, which help attract participation and maintain a healthy ecosystem that encourages accountability in both subjects and practice of reporting.

3. FROM FINDING ANSWERS TO FINDING QUESTIONS

Much of the database research to date has focused on answering questions. For journalism, however, finding interesting questions to ask is often more important. How do we define interestingness? Where do we come up with interesting questions? To gain some insights, we start with two news excerpts as examples:

The water at American beaches was seriously polluted ... with the number of closing and advisory days at ocean, bay and Great Lakes beaches reaching more than 20,000 for the fourth consecutive year, according to the 19th annual beach water quality report released today by the Natural Resources Defense Council (NRDC). ... [10]

... Hopkins County also maintained the lowest monthly jobless rate in the immediate eight-county region for the 29th consecutive month. ... [11]

Like the two excerpts above, many news stories contain factual statements citing statistics that highlight their newsworthiness or support their claims. Oftentimes, these statements are essentially English descriptions of queries and answers over structured datasets in the public domain.

What if, for each such statement in news stories, we get a pointer to the relevant data source as well as a query (say, in SQL) whose answer over the data source would support the statement?

With such information, we can turn stories *live*. We are used to static stories that are valid at particular points in time. But now,

our system can continuously evaluate the query as the data source is updated, and alert us when its answer changes. In the beach water quality example, a continuous query would be able to monitor the alarming condition automatically year after year. If more detailed (e.g., daily) data are available, we can tweak the query to make monitoring more proactive—instead of waiting for an annual report, it would alert us as soon as the number of closing days this year has reached the threshold. Thus, the query in the original story lives on, and serves as a lead for follow-ups.

We will be able to make stories *multiply*. The original story may choose to focus on a particular time, location, entity, or way of looking at data. But given the data source and the query, we can generalize the query as a parameterized template, and try other instantiations of it on the data to see if they lead to other stories. In the jobless rate example, an interested user may instantiate another query to compare her own county of residence against its neighbors. Note that the original query requires more skills than might appear at first glance: it searches for a Pareto-optimal point (x, y) to report, where x is the number of neighboring counties and y is the number of consecutive months. By enabling a story to multiply, we facilitate reuse of investigative efforts devoted to the story, thereby alleviating the lack of expertise, especially at smaller local news organizations.

We will be able to *fact-check* stories quickly. Fact-checking exposes misinformation by politicians, corporations, and special-interest groups, and guards against errors and shady practices in reporting. Consider the following example from FactCheckED.org [12], a project of the Annenberg Public Policy Center. During a Republican presidential candidates' debate in 2007, Rudy Giuliani claimed that adoptions went up 65 to 70 percent in the New York City when he was the mayor. The city's Administration for Children's Services (ACS), established by Giuliani in 1996, made a similar claim by comparing the total number of adoptions during 1996-2001 to that during 1990-1995.

If we were given the data source and the query associated with the claim above, our system can simply run the query and compare its result against the claim. You may be surprised (or perhaps not so) to find that many claims exposed by FactCheckED.org cannot even pass such simple checks. This example, however, requires more effort. According to FactCheckED.org, the underlying adoption data, when broken down by year, actually show that adoption began to slow down in 1998, a trend that continued through 2006. Lumping data together into the periods of 1990-1995 and 1996-2001 masks this trend.

Even when simple automatic fact-checking fails, making sources and queries available goes a long way in helping readers uncover subtle issues such as the one above. When reading stories (or research papers), we often find ourselves wondering why authors have chosen to show data in a particular way, and wishing that we get to ask questions differently. In reality, most of us rarely fact-check because of its high overhead—we need to identify the data sources, learn their schema, and write queries from scratch. By making sources and queries available for investigation, we can significantly increase the crowd's participation in fact-checking to help us ensure accountability.

We have discussed three useful tools—making stories live, making stories multiply, and fact-checking stories—all based on the assumption of having sources and queries to support claims. Naturally, the next question is how to get such information. We could ask makers of claims to provide this information (akin to

requiring data and methods for scientific papers), but this approach does not always work. These people may no longer be available, they may have obtained the answers by hand, or they may have motives to withhold that information. Instead, can our system help us identify the data source and reverse-engineer the query associated with a claim?

This problem seems to entail solving the natural language querying problem, which several research and productization efforts attempted in the past but has not caught on in practice. We believe, however, that two new approaches will give us extra leverage. First, we have the text not only for the query, but also for its answer. Evaluating a candidate query on a candidate dataset and comparing the answer can serve as a powerful confirmation. Second, and perhaps more importantly, as more people use our tools on stories, our system can build up, over time, a library containing a wealth of information about data sources, queries and answers, as well as how they are used in actual stories. Suggestions for relevant data sources may come from stories on similar topics. Reverse engineering of queries can benefit from seeing how similar texts have been translated.

This library also leads us to the interesting possibility of building a reporter's black box. An investigative piece may involve hundreds of hand-crafted queries on a structured database. Apprentices of investigative journalism face a steep learning curve to write interesting queries. In fact, the majority of these queries seem to conform to some standard patterns—grouping, aggregation, ranking, looking for outliers, checking for missing values, etc. A reporter's black box will be a tool that automatically runs all sensible instantiations of "standard" query templates on a database. For databases that are updated, the black box will automatically monitor them by evaluating the queries in a continuous fashion. The library of datasets and queries maintained by our system will help us discover and maintain collections of interesting query templates. It will also help us find patterns across datasets (or those with particular schema elements), allowing us to "seed" template collections for new datasets.

The number of interesting query templates for a dataset may be large, and the number of instantiations will be even larger, since a parameter can potentially take on any value in the dataset. When the reporter's black box runs on a dataset, it should present query-answer pairs in order of their newsworthiness, which helps journalists focus their efforts. Ranking criteria may incorporate generic ones such as query length (compact queries are more compelling) or template-specific ones such as answer robustness (answers that change with small perturbations to query parameters are less compelling). The library of datasets and queries maintained by our system is also useful. For example, queries with templates that have been used by many high-impact stories probably should be ranked higher, especially if their answers have not appeared in old stories based on the same templates.

Running a large number of queries and monitoring tasks *en masse* poses interesting system and algorithmic challenges. Cloud parallelization helps. Techniques in multi-query optimization and scalable continuous query processing are applicable. However, queries in the reporter's black box can get quite complex (if you have trouble motivating skyline queries, look here), which complicates shared processing. On the other hand, more sharing arises from the fact that many queries are instantiated from the same template. The need to produce ranked query-answer pairs also presents unique challenges and opportunities. Instead of devoting an equal amount of computational resources to each query, we would give

priority to queries that are more likely to yield high-ranking query-answer pairs.

Interestingly, there is one area of news where a specialized reporter's black box has been immensely successful—sports. It is amazing how commentaries of the form "player X is the second since year Y to record, as a reserve, at least α points, β rebounds, γ assists, and δ blocks in a game" can be generated seemingly instantaneously. Replicating this success for investigative journalism is difficult. While sports statistics attract tremendous interests and money, public interest journalism remains cash-strapped, and has to deal with a wider range of domains, smarter "adversaries," more diverse and less accurate data sources, and larger data volumes. The ideas presented in this section will hopefully help combat these challenges, by providing efficient, easy-to-use computational tools that aid journalists and citizens in their collaboration to ensure accountability, and by creating a positive feedback cycle where participation helps improve the effectiveness of these tools.

4. CONCLUSION

In this short paper, we have outlined our vision for a system to support collaborative investigative journalism. We have focused on several features of the system to highlight a few important database research challenges. As the Chinese saying goes, we are "throwing a brick to attract a jade"—there are many more interesting problems, both inside and outside the realm of database research: privacy, trust and authority, data mining, information retrieval, speech and vision, visualization, etc.

The need to realize this vision is already apparent. With the movement towards accountability and transparency, the amount of data available to the public is ever increasing. But at the same time, the ability to work with data for public interest journalism remains limited to a small number of reporters. Computation may be the key to bridge this divide and to preserve journalism's watchdog tradition. We hope to motivate you, both intellectually and civically, to join us in working on computational journalism.

Acknowledgments The authors would like to thank all participants of the 2009 workshop on Accountability through Algorithms: Developing the Field of Computational Journalism, organized by James T. Hamilton and Fred Turner, and hosted by the Center for Advanced Study in the Behavioral Sciences at Stanford.

C.L. and J.Y. would also like to thank HP Labs in Beijing, China for hosting their visits in the summer of 2010, during which some ideas in this paper were discussed.

5. REFERENCES

- [1] Pew Research Center's Project for Excellence in Journalism. "The state of the news media." March 15, 2010.
- [2] American Society of News Editors. "Decline in newsroom jobs slows." April 11, 2010.
- [3] Andersen. "Four crowdsourcing lessons from the Guardian's (spectacular) expenses-scandal experiment." Nieman Journalism Lab. June 23, 2009.
- [4] Doan et al. "Community information management." *IEEE Data Engineering Bulletin*, 29(1), 2006.
- [5] Jeffrey, Franklin, and Halevy. "Pay-as-you-go user feedback for dataspaces systems." *SIGMOD* 2008.
- [6] Welsh and Smith. "Highest crime rate in L.A.? No, just an LAPD map glitch." *The Los Angeles Times*. April 5, 2009.
- [7] Dalvi, Rá, and Suciu. "Probabilistic databases: diamonds in the dirt." *CACM*, vol. 52, 2009.
- [8] Widom. "Trio: a system for data, uncertainty, and lineage." In Aggarwal, editor, *Managing and Mining Uncertain Data*, Springer, 2009.
- [9] Madden et al. "TinyDB: an acquisitional query processing system for sensor networks." *TODS*, 30(1), 2005.
- [10] Natural Resources Defense Council. "Beach closing days nationwide top 20,000 for fourth consecutive year." July 29, 2009.
- [11] Bruce Alsobrook. "Local unemployment rate best in eight-county area for 29th straight month." *The Sulphur Springs News-Telegram*. September 22, 2010.
- [12] FactCheckED.org. "Dubious adoption data." June 6, 2007.

Beyond the Foreclosure Crisis Housing Strategy for Cleveland's Future

Brian Mikelbank, Director
Charlie Post, Research Associate
Ivan Maric, Research Assistant
Tom Bier, Executive in Residence

Center for Housing Research and Policy
Maxine Goodman Levin College of Urban Affairs
Cleveland State University

September 2008

Introduction

The Center for Housing Research and Policy at the Maxine Goodman Levin College of Urban Affairs, Cleveland State University, has completed and details herein three related analysis tasks regarding the housing market of Cleveland, OH. These are meant to provide information at two levels. First, all analyses contain a citywide component, so that the performance of any particular indicator can be seen for the city as a whole, and most often across time. Second, we provide city ward level detail so that the spatial detail in the data can be assessed.

The first task documents the lending activity in the City of Cleveland of all the major lending institutions. We rely on Cuyahoga County records of Mortgages and Deeds from 2003 through the third quarter of 2007 to provide detail on approximately 47,000 mortgages of single-family properties in the City of Cleveland. This was the time of the boom in both home lending overall and sub-prime lending in particular. The analysis shows a consistent range of low, no, or negative equity lending, making the resultant patterns of foreclosure less than surprising. While the whole city suffered from this type of activity, the issue is clearly more pressing in distinct areas of the City, and is more prevalent among distinct lenders and types of lenders.

The second task details price and volume trends of single family houses among three distinct housing submarkets. Sales across these markets are analyzed for Cuyahoga County, the suburbs of Cuyahoga County, the City of Cleveland and the City of Cleveland's twenty-one wards. We find that the submarket impacted directly by foreclosure is in fact in crisis – volumes are up, and price declines are substantial. However, for the portion of the market that has avoided the foreclosure process, prices are relatively stable. The volume of this latter market is decreasing, but this could well be a voluntary act on the part of potential sellers, holding out until the market improves. We also analyze activity among the recently/newly constructed market, but only in aggregate, since the volume of market activity is low.

The third task investigates two areas of potential trouble in the housing market. The first is quick-resales, where a house transacts twice over a short period of time, and at an unusual price increase. The second area is that of multiple-transactions, where we identify entities whose pure volume of activity in the city could be cause for concern. Both of these cases serve as an inroad to identifying market activity that could have negative externality impacts on neighborhoods. For example, flipping a house in the absence of any meaningful additional investment in the property provides no neighborhood benefit. Similarly, an institution that is simply in the business of buying and selling large volumes of properties across the city (or region) likely has no tie to the neighborhoods in which they operate – no benefit of the stability of longer term residence, no pride of ownership, etc.

In sum, the report lays the foundation of the indicators necessary to signal the return of a healthy Cleveland housing market: equity in home purchases, and thus investment in the city's neighborhoods; a market balance shifted in favor of homes free from the direct impacts of foreclosure; and market activity that is indicative of home ownership, responsible rental activity, or meaningful redevelopment.

Task 1: Lending, Equity, and Value
Table 1: Home Equity in Single Family Homes

Background. Table 1 shows the equity position of single-family home purchases in each ward since 2003. It compares the loan amount for the purchase to the value of the home (as assigned by the County Auditor) and the actual sales price. The group of columns on the left hand side of Table 1 shows information about the median home loan value, versus the median home value, as determined by the County Auditor. Ideally, the loan-to-value ratio would be less than 100%, meaning that the loan amount was less than the house value – this would be an indicator of equity. On the right hand side of Table 1, we compare the loan amount to the actual sales price. A loan-to-price ratio of less than 100% would again reveal equity in the home purchase. Where homebuyers have larger down payments and more equity, they are more invested in the neighborhood, and potentially have a stronger financial foundation from which to enter into home ownership and its related costs.

Analysis. While the down payment in the City as whole was 5% (right hand side columns – a loan-to-price ratio of 95%), a more realistic equity position is shown to be negative 5.6% (left hand side of the table – a loan-to-value ratio of 105.6%). The difference between the auditor value and actual sale price shows how the process of buying a house and getting a mortgage was potentially manipulated during this time period.

There also appears to have been a difference between the East Side of Cleveland and the West Side. In the East Side, the auditor-based equity position was a staggering negative 15.1% (a L/V ratio of 115.1%), while the loan-to-price equity measure was positive 5%. In the West side, the auditor based measure was exactly 1 (a ratio of 100%) and the later was 3%.

There were a couple of wards where the auditor-based equity position was robust (such as Wards 16 and 21), but we find most of the wards had a substantial amount of lending that potentially left the borrowers with negative equity, according to the auditor based values. In fact, in Ward 7, we had a case where the loans were over 40% greater than the value of the properties used as collateral. In six other wards the negative equity was over 20%.

The price analysis looks better, with equity values nearly reaching 10% in some wards, but median prices being so consistently in excess of the auditor value is suspicious, at best. Further, even if the price data were taken at face-value, equity positions were rarely greater than 5%, and with median prices around \$85,000 this would still represent a relatively small investment in the house. There is a clear dichotomy between the traditionally desired value of a 20% down-payment, and the glaring reality of the equity challenges revealed here.

Summary. Across the city 1/3 of the wards had a situation where the loans were at least 20% greater than the auditor assigned value of the property. Traditionally, one might look for *positive* equity, or a down payment of 20% of the price/value of the property, not *negative* 20% as was revealed here. Low (or negative) equity home

purchasers literally and figuratively are less invested in their neighborhoods, making “walk-aways” and abandonment a distinct possibility.

The reference point of 20% equity is the basis of the categories used in the following tables.

Next Steps. Certainly the ward-level analyses gloss over the detail present within each ward. There may be specific sub-ward/neighborhood level analyses that are of particular interest to Council. Alternatively, we could return to the parcel level data and identify clusters of high and low equity transactions. With additional data from the County Recorder, these analyses/tables could be updated through the first half of 2008 and we could identify changes in market equity by ward or in locations of specific interest. An improvement in the equity position of home purchases would signal an improvement in the market. The pervasiveness of such low equity purchases highlights the need for financial counseling for potential homeowners.

Task 1: Lending, Equity, and Value

Table 2: Loan to Value Ratio, Single Family Homes

Background. For the ease of interpretation we have defined 4 groupings of loans based on their loan-to-value ratio. This provides more detail than the median ratio, which was presented in Table 1. The groupings could be thought of as ranging from extremely risky (Loan-to-Value ratio of more than 120%) to traditional/low-risk loans (with a L/V ratio of less than 80%). From this table one can assess the overall equity distribution of loans by ward, by East or West side of the city, or the City of Cleveland as a whole.

Analysis. While the West Side of Cleveland has a ‘balanced’ distribution of groupings, the East Side shows a clear tendency towards ‘high-risk’ lending. In the West Side, each group had over 20% and less than 30% of all observations – while on the East Side over 45% of mortgages belonged to the ‘high-risk’ grouping where the loan value exceeded the house value by at least 20%. The distribution is skewed, somewhat, towards the extremes at the ward and City level.

One third of wards had over 50% of mortgages belonging to the ‘high-risk’ group, with an additional 2 wards missing that designation by less than 1%. On the other hand Ward 21 was spectacularly balanced towards ‘low-risk’ mortgages, with *only* 12.5% belonging to the ‘high-risk’ grouping. Wards 16 and 20 also had a nice balance towards ‘low-risk’ mortgages. On the East side only Wards 1 and 11 did not have over 40% of their mortgages belong in the ‘high-risk’ class.

Summary. The distribution of equity values by category was not uniform across the City. Ideally, values in the 80% column would dominate each ward, but that was not the case. In no ward did these favorable loans exceed 35% of the total, meaning that the vast majority of loans across the city, across the years since 2003 have had a higher risk than desired. A high loan to value ratio means low equity, and potentially high neighborhood volatility, especially in the context of a weak regional housing market and a weak regional economy.

Next Steps. The use of these categories will make it easy to observe trends in the ratios over time. It is a potential indicator of the financial investment residents have in their homes. Perhaps the most telling of these is the proportion of all loans in the high-risk category.

Task 1: Lending, Equity, and Value

Table 3: Loan to Value Ratio, Single Family Homes (mortgage bankers)

Table 4: Loan to Value Ratio, Single Family Homes (non-mortgage bankers)

Background. While Table 2 did show many variations across wards, it hid some of the variation due to the types of institutions originating the loans. In Table 3 we re-created the logic behind Table 2, but using *only* the cases where the loan originator was a mortgage banker. Table 4 provides the same information, but for cases where the lender is not a mortgage banker. Since the practices of non-traditional lenders have come under scrutiny, the balance of traditional and non-traditional lenders in a neighborhood is of interest, particularly if their lending practices are distinct from one another.

Analysis. Tables 3 and 4 reveal substantial differences in the loan characteristics between the two lender types. For example, Table 3 shows that nearly 40% of mortgage banker loans in Ward 1 were high risk (greater than a 120% loan-value ratio), while Table 4 shows that under 20% of the loans by traditional lenders were high risk. Citywide, only 10.2% of mortgage banker originated loans were low-risk (less than 80% loan-value ratio), while traditional lenders managed to achieve nearly 40% of their loans in that category. More than seven out of every ten (73.1%) mortgage banker loans fell into one of the two negative equity categories.

It should be noted though, that high risk loans were made by both types of institution, and that low equity loans were a necessary component of moving a customer into home ownership if cash reserves were low, or even nonexistent. It is possible that the mortgage bankers served that population more frequently, resulting in lower equity.

The implication for a neighborhood housing market, though, is that mortgage bankers more frequently made loans that negatively impact residential stability. As an indicator of market health, it would be advantageous to see a greater market share develop among traditional lenders across the city.

Summary. The distribution of 'high-risk' lending originated by mortgage bankers was overwhelming, especially on the East Side. Other than Ward 1 (which still had nearly 40% risky loans), the East Side was dominated by the negative equity lending practices of the mortgage bankers. Less than 8% of East Side mortgages were 'low-risk' loans (e.g. 20% positive equity).

The distribution of risk class of non-mortgage bankers, shows that their lending practices compared favorably to non-traditional lenders, although the proportion of 'high-risk' mortgages was still too high. Only a one ward (21) had single digit high-risk lending over the time period, and the East Side was nearly at 30%, even among traditional lenders.

Next Steps. Given the difference between the equity distributions of the two different types of lenders, it would be of specific interest to encourage all potential homeowners explore all of their financial options when shopping for a mortgage. Anything the City could do to facilitate a greater presence of traditional lenders in its neighborhoods might increase their market share, and increase homeowner investment, and decrease the kind of lending practices that could bring instability (foreclosure, abandonment, etc.).

As time passes, it would be of interest to monitor the performance of these lending categories to see the degree to which the tightening of the lending market has influenced the equity position of their loans. If part of that higher standard is a stricter equity/down payment requirement, the impact on neighborhoods would be two-fold. First, the loans that are made would be more stable due to homeowners' increased investment in their properties. Second, the number of loans could be less since potential homeowners that don't meet the new requirements could be left out of the market. This could result in an oversupply of houses on the market, which could dampen any market recovery.

Task 1: Lending, Equity, and Value
Table 5: Home Equity, Individual Lenders

Background. The analysis of the mortgage lending practices across wards and institution types has shown there to be great variation in equity across each category studied. In this table, we disaggregate those results to identify the top 5 individual institutions in each ward. We identify the type of institution (CB = Commercial Bank, MB = Mortgage Bank, SL = Savings and Loan, OT = Other), and summary information about its lending profile by ward. This information could prove useful in seeing the major lending actors in a specific ward, and also in comparing lender activity across wards where they are a prominent lender.

Analysis. The first thing that emerges from this table is the negative effect of Argent Mortgage Co. on the whole City market. It was in the 'Top 5' institutions in all 21 wards. In each ward it had median Loan-to-Value ratio that was over 100% (indicating negative equity), and in many cases it was over 140%. Only in Wards 1, 16, 20, and 21 was this ratio less than 120%. The only other institutions that show any similar affect on the Cleveland mortgage market were the People's Choice Home Loans and New Century Mortgage Corporation. Even their combined effect, though, was less than Argent's alone.

An interesting pattern in the data is how the L/V ratio is typically well above 120% for these three institutions, while in most cases the L/P ratio was steadily hovering near 90%-95%. One might initially write this off as a quirk of the data, but it is notable that this pattern does not persist with traditional lenders. This points to the reality that among those three lenders it was common for houses to sell well beyond their auditor determined market value, and to have a loan for at least 90% of that price. That loan amount, then, is typically in excess of the auditor market value, resulting in the negative equity findings described here.

The institutions that were in the observed in the 'Top 5' and which originated mortgages with positive equity were: National City Bank; Third Federal S&L; KeyBank; Charter One Bank; and Huntington Bank. National City Bank was in the 19 of 21 wards, and in all but 2 of those wards, it had a positive median equity position on its loans. Only in Ward 12 was the median L/V ratio meaningfully negative (in Ward 5, the ratio was 100.6%).

The next most significant institution was the Third Federal S&L. It was in the 'Top 5' in 18 of 21 wards, and it had median positive equity in all but 4 wards. Our data indicate that only in Ward 8 was this median L/V ratio meaningfully negative.

KeyBank was a 'Top 5' institution in 9 of 21 wards and it had median positive equity in all but 2 wards. Our data indicate that only in Ward 9 was this median L/V ratio meaningfully negative (at 108.8%). What is interesting is that it shows up as predominantly an East Side lender. And, in the column describing the 'down-payment' it shows a minimum down-payments of around 1% (L/P ratio of 99%).

Summary. The 'Top 5' institutions data reinforce how the most active mortgage bankers differed from the top non-mortgage bankers throughout the city. The mortgage bankers had median equity positions that were in the negative 20%-50% range, but with down-payments of positive 5%-10%. The traditional institutions showed positive equity, but with a down-payment that was in the 1%-6% range. In other words, the mortgage bankers had a consistent discrepancy between the transaction prices and market values related to their lending activity.

Next Steps. It will be important to monitor future lending activity, both in terms of institutional volume and institutional characteristics. Doing so would alert neighborhoods to the emergence of new lenders (and their characteristics), the disappearance of lenders, and how market entry and exit influences the balance of equity and lending activity.

Task 2: Single-Family Sales: Three-Segment Market Indicator

Background. We argue that treating the single family sales market as one single market is no longer appropriate due to the direct impact the foreclosure process is having on the local market. The purpose of these charts, therefore, is to highlight trends in single-family sales counts and median prices for two separate parts of the market: those properties which have been directly impacted by foreclosure activity, and those properties which have not been directly impacted.

We originally planned an additional market segment (New construction built since 1990). However, the number of sales in this category turned out to be so low for small geographic areas that we decided to track these sales only for the City of Cleveland as a whole, and to break it out into the not impacted and directly impacted categories.

This indicator set can provide useful signs that the market is improving. These might include: the number of not impacted sales increases; the number of directly impacted sales declines; prices stabilize or increase in either portion of the market. The newly constructed resale activity is included to see if this market segment is faring any differently than existing construction more broadly. Any distinction between the newly constructed market and the longer-term existing market could provide policy direction.

The charts include only single-family sales of at least \$500, and include only the standard deed types that the Center for Housing Research & Policy have traditionally used to represent true market activity. We have produced the charts in hard copy format, but the data remains in electronic form only. For this analysis, directly impacted means one or more of the following is true:

- (a) The current sale is a sheriff sale
- (b) The property was auctioned at a sheriff sale within the past two years
- (c) The property was involved in a foreclosure filing within the past two years

“Not impacted” means that none of the above conditions listed is true. We are concerned here with the *direct* impact. It is possible that sales near foreclosed properties experience *indirect* proximity effects of surrounding foreclosure activity.

Information is displayed for both sales volume and median sale prices. These are shown for the following geographic areas: each of the 21 Cleveland city wards, the City of Cleveland as a whole, Cleveland east side and Cleveland west side, Cuyahoga County as a whole, the suburbs as a whole, and the suburbs east and suburbs west. While ward-specific detail is provided for a more limited audience, we suspect the charts detailing the more aggregate geographies will be of broader interest. For example, one might want to compare individual wards to city performance overall, or compare overall city performance with that of the suburbs.

The directly impacted sales are consistently indicated by the dashed lines, while the sales not directly impacted are indicated by the solid lines. Unfortunately, it is sometimes the case that the number of sales is so low that a median price statistic is not reliable. If the number of sales is under 20, we have indicated that the median is

not reliable by displaying an open circle for the data point in the median sale price chart. For example, in Ward 3, by first quarter 2007, the number of not impacted sales is under 10, and the median price is highly variable. The greater the number of observations, the more reliable is the median price measure.

Analysis. Perhaps the trend of greatest concern in the data is shown in the chart, "Cleveland Single Family Sales Counts, Existing Homes." Starting in 2005, the number of directly-impacted sales started to increase dramatically, while by the end of 2005, the number of not impacted sales started to drop off in equally dramatic fashion. Through the end of 2007, this trend persisted, with only a slight drop off in impacted sales through the first quarter of 2008. While not impacted sales are distributed relatively evenly between the East and West sides (with just over 100 per quarter on East, and approximately 150 on the West), directly impacted sales are more prevalent on the East side, with over 700 sales per quarter since the beginning of 2007.

In terms of sales volume, the segment of the market that has been directly impacted by foreclosures has been accelerating, while the remainder of the market continues to stall. The Cleveland market has become dominated by these directly impacted transactions, and these are likely governed by an entirely different set of decision rules than not impacted sales are. This directly impacted portion of the market is fundamentally different than the traditional market, and it has become the dominant force in Cleveland.

However, we do not find the entire market to be in a price free-fall. The charts "Cleveland Single Family Median Sale Prices, Existing Homes," and the component charts for the East and West side show that the market segmentation noted above also extends to sales prices. Citywide, directly impacted prices had hovered between \$40,000 and \$50,000 through the end of 2006, at which point prices began a steady decrease through the first quarter of 2008.

Prices of not impacted properties however, have not seen the same dramatic declines. Citywide, median sales prices of not impacted properties have been holding at nearly \$80,000 since the beginning of 2003. On the West side, prices have held at above that mark, and on the East side, have been in a gradual decline since the beginning of 2007, with the median price dipping below \$70,000 in the first quarter of 2008.

However, since the market is increasingly dominated by the directly impacted activity, and since the prices associated with that activity are substantially lower and on a decreasing trend, aggregate reports of city housing activity tend to be misleading. If they fail to acknowledge these two very different types of market activity, reports will be negatively biased. For example, despite volumes that are indicative of a troubled market, the not impacted portion of that market has not experienced the dramatic price decreases that many suspect. For sales that have not been part of the foreclosure process, prices are not markedly different than they were in 2003.

Naturally, the market should be gaining in value, but the finding that the traditional portion of the market has been relatively price-stable is an important one, especially for homeowners who are able to steer clear of foreclosure. The concern is volume – that the not impacted portion of the market is increasingly small, and has not yet

shown signs of growth.

For housing constructed since 1990 ("Cleveland Single Family Median Sale Prices, New Construction (Built since 1990)," the price story is similar. For homes directly impacted by foreclosure, prices have been around \$80,000 since early 2006. For those not directly impacted, prices have been fluctuating around \$160,000 over that same time period. Sales volume, again, provides additional insight. Sales volumes of both impacted and not impacted are relatively low given the overall distribution of units constructed since 1990. Also, impacted and not impacted sales have been much more balanced compared to the rest of the market, especially since 2007. The finding that the newer construction market has had higher prices and lower sales volume points to a greater stability in this market than is present in the market more generally.

Summary. The simple indicators of overall market volume and average market prices are no longer up to the task of meaningfully monitoring Cleveland's housing market. Distinct submarkets have emerged and are far more useful in understanding the current dynamics in the market. The submarket that has been directly impacted by foreclosure is in fact cause for concern. Prices are decreasing, and the volume of this submarket is on the rise. Equally troublesome to this rise in activity is that it is increasing proportionally to "healthy" market activity. As the volume of sales not impacted by foreclosure decreases, significant downward pressure on average and median prices is the result. When these aggregate indicators subsequently get reported in the press, the process feeds upon itself. However, despite decreasing volume in the not impacted segment of the market, we found prices to be much more stable than one would otherwise anticipate. Thus, the silver lining of these analyses is that although the market is smaller, its performance is surprisingly stable.

Next Steps. When the market starts to improve, we would expect a more favorable balance of the impacted and not impacted submarkets. In addition to a better balance, volumes of not impacted sales need to increase across the city. With the higher prices obtained from not impacted properties, the overall prices will also start to increase, which will attract more activity back to the market. This could trigger a potentially severe over supply issue, though. As confidence returns to the market, those households waiting to sell will also return. It is likely that given the current market perceptions, many are waiting to sell their houses until the market "turns around." With the return of that activity, along with the supply of housing coming out of the foreclosure process, and the healthy supply already on the market, price increases will likely lag the return of healthy market activity as prices adjust to, but eventually recover from, over supply.

Task 3: Quick Resale and Multiple Transactions

Background. Another unhealthy market characteristic that has developed throughout much of the county is property churning, where homes are repeatedly selling between market participants, generally with no plans to occupy or materially improve the property. We focus on two primary components of this activity: the timing of transactions (what we call “quick resale”) and the frequency of participation in this type of activity (what we call multiple transaction activity).

Property churning is problematic for local housing markets. Ownership becomes more difficult to track, making it more difficult for cities to hold owners responsible for their properties. It perpetuates a string of owners not truly invested in the house, the neighborhood, or the city. Thus, we track quick resale and multiple transaction activity for single-family, two-family, and three-family houses, as well as condominiums.

Names have been standardized as much as possible to consolidate all of the various spellings of the buyer and seller names. For example, “Cleveland Housing Network”, “Cle Hsg Ntwrk”, and “Cleve H Network” all are consolidated under one name for the purposes of these analyses. However, there are undoubtedly some which slipped through this process. For every quarterly run of these data, we will again look for additional name combinations for standardization, and thus over time more and more participants acting under slightly different names will get standardized into our quick resale and multiple transaction databases.

For quick resale activity, we identify a listing of entities which have bought and sold properties very quickly (compared to traditional market transactions) and at an unusually high markup. Being on this list is not evidence that the owners are doing something untoward, because there are many legitimate reasons why the property could quickly appreciate. However, this list does provide a starting point for considering who might be engaged in “flipping” a house without adding any true value to it.

A quick resale is defined here as a resale of a property within 90 days of the previous sale, and at a resale price which is at least 25 percent above the previous price. In many cases, the initial purchase price is very low, sometimes under \$1,000, and the resale price is also quite low. However, if the resale took place within 90 days, and the price increase was at least 25 percent, then it was included in the analysis.

To provide a sense of the typical price range for the resale for each seller, the median price across all resales by that seller is included in the table. This allows one to evaluate the dollar amount represented by the increase in the quick resale price. For example, a \$125,000 median resale price means that the seller made at least \$25,000 on the typical resale, while a \$12,500 median sale price means the seller made at least \$2,500 on the typical resale.

We show annual totals for the years 2006 and 2007, and then the data for 2008 are shown by quarter. Individual listings for sellers are provided if they had at least two quick resales in the quarter. The total includes many single-resale sellers which are

not listed separately.

For the multiple transactions tables, we include activity as both the buyer and the seller, which means that the sum of the “buys” and “sales” equals twice the number of actual transactions. However, looking across any row in the table, it is usually the case that for any individual entity, it is either a buyer or a seller in any given transaction. There are some notable exceptions, for example Cleveland Housing Network, which often is both the buyer and the seller. If a transactor is both a buyer and a seller in a transaction, then the column “# Buys” is incremented by one, as is the column “# Sales”, and the cell in the appropriate ward is incremented by two, representing both ends of the transaction. In a more typical transaction, either the “# Buys” or the “# Sales” column is incremented by one, as is the appropriate ward cell.

In order to make the ward totals more useful and intuitive, they have been converted to represent the total number of transactions. Thus, even though each produces both a buy and a sale, the ward totals count the exchange only once. These can be found in the bottom row, labeled “TOTAL single-count transfers”.

Entities are listed separately in the tables if they meet a threshold number of total buys + sales for the period. For single quarters of the year, the threshold is five. For year to date analyses, the threshold is increased by five each quarter, so the thresholds are ten for the second quarter, fifteen for the third quarter, and twenty for the fourth quarter. Twenty is also the threshold for tables which represent entire years of data (2006 and 2007).

The median purchase and/or sales prices are also included in the table for each entity. Separate “non-zero” counts are included because for the median prices, all transfers at zero dollar value are removed. Many transfers take place at zero value, and these transfers would distort the median values. We provide annual data for 2006 and 2007, with quarterly detail for the first half of 2008.

Extra detail for the most active buyers/sellers is also provided in two additional tables produced for 2007. One is represented by the title, “Who sold to buyers of 50 or more properties in 2007?” All entities which bought at least 50 properties during the time period are listed across the top. Down the first column are listed sellers of properties to this subset of big buyers. The other type of summary table is represented by the title, “Who bought from sellers of 50 or more properties in 2007?” In these tables, all entities which sold at least 50 properties during the time period are listed across the top. Down the first column are listed buyers of properties from this subset of big sellers.

Analysis.

Quick Resale: The tables provided show the changing nature of this portion of the market, and allow identification of the major and emerging players in it. For example, it is noteworthy that when we conducted this type of analysis many years ago, the resale prices in general were much higher for the general population of quick resellers. Even as recently as 2006, the median resale price was \$82,000. In 2007 the median resale dropped to below \$37,000. The first quarter of 2008 saw a median price of only \$6,100, and by the second quarter of 2008 this had dropped by nearly another

thousand dollars to \$5,150. Clearly the nature of the activity in this segment of the market has changed over the last few years.

This change is underscored by looking at the individual actors. For example, the current activities of Destiny Ventures have been well publicized, but they had only three quick resales in 2006, with a median sales price of \$4,000. In 2007, they increased to 35 resales, and their median resale price decreased by 50%, to \$2,000. Through only the second quarter of 2008 they have had 53 quick resales, and the median resale price was less than \$2,000 in both quarters.

Similarly, Blue Spruce did not even meet the criteria to be on the quick resale list in 2006, had only four quick resales in 2007 (with a median resale price of \$1,250), but has had 48 quick resales through the second quarter of 2008 (with median resale prices of \$1,000 in both 2008 quarters).

Note that the table for the second quarter of 2008 provides current data for that quarter, but also the first quarter's activity and a year-to-date tally. This would help identify changes in participant behavior. For example, REO Nationwide LLC was party to 10 transactions in Q1, but 16 in Q2. Lender USA had one quick resale transaction in Q1 but four in the second quarter. This quarterly information will prove useful in identifying new players in this market.

Each quick resale table also provides ward level detail. One could scan across any line and see the location of activity of a particular market participant by ward, or could look to the bottom of the table to see total quick resale activity across all market participants for the current quarter and year-to-date.

Multiple Transactions. Scanning the 2006 and 2007 annual tables, at least three types of activity stand out. First is the situation in which many mortgage holders find themselves. Consider, Deutsche Bank National, for example, that purchased 314 homes at a median price of \$40,000, and sold 229 homes at a median price of \$22,500. This is likely the result of taking back properties via Sheriff sale auctions, and then selling them for whatever possible given the current market conditions, seemingly losing a significant amount of money on each transaction. The situation is similar for many of the lenders and mortgage holders one might recognize looking down the "Transactor" column of the multiple transaction tables.

A second variety of activity is evidenced, for example, by York Management. They purchased 25 properties at a median price of \$28,250, and sold 25 properties at a median price of \$90,000, seemingly profiting from its transactions. It should be explicitly stated that we don't have information on any improvements that were undertaken on any of the transactions detailed here. So, for example, York Management could be investing heavily in the houses it sells. They could be purchasing only inexpensive properties while selling off more expensive properties from its portfolio. If transactions were over a very short period of time, with a high level of profit, the activity would show up in the quick resale tables. Timing that is different, profits that fall below the threshold, or some combination of these could keep that activity off of the quick resale list, but simply the volume of transactions land it on the multiple transaction list. Thus, appearance on this list is no sort of

indictment against market participants. The purpose is to highlight the activity, price points, and location of the market's most active players.

Third, there are the transactors operating at the bottom of the market, where median purchase and sales prices are extremely low. From the 2007 table, a few stand out: Econohomes, Destiny Ventures, REO Nationwide LLC, New Day Development, Blue Spruce, and Jeffery Clark. There are many other market participants acting in the manner, but on a much smaller scale.

Each type of multiple transaction activity listed above should be of interest to the City of Cleveland generally, but ward leaders specifically. For bank activity, sales tend to be at a low price point, and have the potential to pull down surrounding neighborhood house values. Additionally, volumes are high for both sales and purchases. This creates the reality where banks are increasingly becoming neighborhood "residents" without any of the traditional beneficial homeownership ties to the local community.

Identifying multiple transactors with large price differentials will bring to light entities either working diligently to improve the neighborhood, or working diligently to fleece it of any remaining resources. Both types of activity should generate neighborhood leadership interest.

The final group of actors is simply churning a high volume of low quality properties at modest per-property profits. They seem to be providing no community benefit, but only perpetuating a cycle of property delinquency and abandonment.

The final two tables show an aggregate level of property transactions for the city as a whole. The "Who bought from sellers..." table shows sellers of at least 50 properties across the top of the table and buyers of at least ten properties down the left hand side. The cycle of property sales becomes evident when looking at this web of transactions. For example, Deutsche Bank sold 61 of its properties to Destiny Ventures, who sold 20 of its properties to EconoHomes LLC and an additional 13 homes to Stewardship Fund LP. While Stewardship LP bought no other homes from the top sellers listed, Econohomes LLC also purchased homes from eight of the other top sellers.

Lastly, the table, "Who sold to buyers..." looks at the patterns of acquisition among entities that purchased at least 50 homes in 2007. This provides information as to sellers to the city's largest volume buyers. For example, Destiny Ventures purchased its properties, in declining order, from Deutsche Bank National, The Bank of New York, US Bank, JP Morgan Chase Bank, and Wells Fargo Bank.

Summary. Activity in these tables is quite variable from one ward to the next, and is a sign of potential market trouble. Optimistically, it would be a sign of a healthy market if developers were able to buy, rehab, and sell for a profit homes in your city. It would signal strong demand, and an efficient place to invest in housing. Unfortunately, knowing the context and price details of the transactions highlighted for this task, we can be not that optimistic. The price points shown here are often too low to represent any type beneficial investment in housing. When prices increase dramatically, there is reasonable skepticism from just the raw data that such as massive redevelopment effort could take place over the short period of time used as a

quick resale parameter. Further, seeing the way in which multiple properties are being transacted among non-local institutions only serves to heighten concern that the type of market activity that is now prevalent, is having, and will continue to have, severe implications for the city's neighborhoods.

Next Steps. Given the availability of the appropriate data, we could link quick resale reports to city issued building permits. This might reveal which quick resales actually involved a change in housing quality, not just a quick turnaround with minimal investment and higher than usual profits. Surely there are true investments taking place within these data, but at this point they can not be distinguished from the data on more questionable activity. We could also produce lists or maps of the properties involved for neighborhood or ward level investigation of potentially troublesome activity. It is clear though, that the nature of the quick resale activity has changed, even over the last few years, to be dominated by the extreme bottom of the market. These houses are turning over quickly, and likely with little or no reinvestment. Everybody loses in that transaction, except for the one exploiting the market to perpetuate a cycle of abandonment at the cost of the city's neighborhoods. Any administrative check-point, such as a point-of-sale inspection, should have great potential to slow this stream of activity.

The multiple transactions data could likewise be mapped at any geography to reveal the spatial domain of these activities. These data shouldn't be viewed as any less troublesome than the quick-resale data. They could be separated by mere days or dollars from the threshold conditions placed on the quick-resale label. The volume of sales between institutions is troublesome in terms of returning properties to productive use.

Beyond Foreclosure: Summary

The purpose of the research provided here is to present (1) a set of baseline indicators, (2) the aggregate geographic distribution of these indicators, and (3) a facility by which together we can investigate parcel-level distributions of these indicators to explore location-specific strategies.

What is apparent from the indicators provided is that

- It isn't simply the volume of lending activity that is important, but the character of that activity, and particularly the equity represented in each transaction.
- Lenders varied in the profile of the loans they originated.
- Lenders varied in the geography over which they operated.
- Simply monitoring sales prices is no longer a useful approach to understanding the city's housing market. Submarkets have developed, hinging on whether or not the property has been directly impacted by foreclosure.
- There is a variety of market activity that continues to operate at a volume that is detrimental to the city's recovery activities, and that are likely feeding the most negative aspects of the segmented market.

We know that we have presented a significant volume of information here, but we felt it necessary to provide a foundation for all of these market indicators for all areas of the city. Building on this information in subsequent quarterly reports, we will know

when the market is changing, and we will have the luxury of basing that knowledge on actual market information rather than guesswork, speculation or anecdotal evidence. This information will be a key input to any policy driven solutions.

We anticipate, however, that the volume of information presented here will generate additional questions, and our agreement leaves time and resources for these questions to be addressed. We look forward to working with Council further on these issues.

Select Findings from
Beyond the Foreclosure Crisis: Housing Strategy for Cleveland's Future
 Prepared by the Urban Center's Housing Initiative
 Maxine Goodman Levin College of Urban Affairs
 Cleveland State University

Loan to Value Analysis: We compare the amount of the home loan to the auditor-reported estimated market value. This loan-to-auditor value ratio is an indicator of equity. For example, a value of 120% reveals the loan amount to be in excess of the value by 20%, representing potential for a negative equity situation.

TABLE 2
Loan-to-Auditor Value Ratio, Single-Family Homes, 2003-2007(Sep.)

Ward	Less Than 80%	Between 80% and 100%	From 100% To 120%	Greater THAN 120%	Number of Mortgages
1	25.7%	18.8%	25.5%	30.0%	4,041
2	19.0%	13.0%	18.5%	49.4%	2,662
3	18.0%	12.2%	20.3%	49.4%	1,801
4	22.9%	14.0%	18.3%	44.8%	997
5	22.6%	23.2%	14.1%	40.1%	694
6	16.6%	15.1%	14.5%	53.9%	761
7	15.1%	15.0%	10.0%	60.0%	889
8	15.6%	13.3%	17.1%	54.1%	1,413
9	17.1%	11.5%	16.2%	55.2%	1,182
10	19.0%	14.6%	20.4%	46.0%	1,777
11	23.0%	17.0%	23.8%	36.2%	2,834
12	14.6%	11.1%	15.2%	59.1%	2,030
13	21.2%	15.3%	14.9%	48.6%	863
14	17.6%	11.8%	13.3%	57.3%	1,407
15	29.7%	20.3%	23.8%	26.2%	3,187
16	30.9%	23.4%	26.5%	19.2%	3,958
17	17.4%	11.7%	14.5%	56.5%	1,498
18	25.6%	15.6%	20.8%	37.9%	2,340
19	27.2%	21.2%	25.9%	25.7%	3,243
20	28.9%	23.6%	27.4%	20.1%	4,871
21	34.8%	26.4%	26.3%	12.5%	4,970
East Side	20.0%	14.9%	19.6%	45.6%	21,343
West Side	28.6%	21.3%	24.1%	26.1%	26,077
Clev TOT	24.7%	18.4%	22.0%	34.9%	47,420

Prepared by: Center for Housing Research and Policy, CSU (8-5-2008)
 Note: Ward 13 is spread over BOTH East Side and the West Side.

Ideally, values in the 80% column would dominate each ward, but that was not the case. In no ward did these favorable loans exceed 35% of the total, meaning that the vast majority of loans across the city, across the years since 2003 have had a higher risk than desired. A high loan to value ratio means low equity, and potentially high neighborhood volatility, especially in the context of a weak regional housing market and a weak regional economy.

Troublesome Market Activity: An unhealthy market characteristic that has developed is property churning, where homes are repeatedly selling between market participants, generally with no plans to occupy or materially improve the property. We focus on two primary components of this activity: the timing and value changes of transactions (what we call “quick resale”) and the frequency of participation in this type of activity in general (what we call **multiple transaction activity**).

Summary Quick-Resale Information

Year	Number of quick resales	Median resale price	Top reseller by volume (#)	Median resale price, top reseller
2006	491	\$82,000	Cresthaven Development (9)	\$114,900
2007	439	\$36,667	Destiny Ventures (35)	\$2,000
2008, Q1	165	\$6,100	Blue Spruce (27)	\$1,000
2008, Q2	234	\$5,150	Destiny Ventures (43)	\$1,750

A “quick resale” occurs when a house sells twice within 90 days, with a minimum price increase of 25%. The summary table above shows the changing nature of the quick resale market. As recently as 2006, the median resale price was \$82,000 and the top reseller had two striking characteristics – only 9 houses resold, and at a price above \$100,000.

By the second quarter of 2008, total quick resale volume had increased substantially, the median price had dropped dramatically, and the volume of individual market actors had also changed. For example, Destiny ventures had only 3 quick resales in 2006, by had over 50 in the first half of 2008. Similarly, Blue Spruce did not even meet the criteria to be on the quick resale list in 2006, had only four quick resales in 2007 (with a median resale price of \$1,250), but has had 48 quick resales through the second quarter of 2008 (with median resale prices of \$1,000 in both 2008 quarters).

Summary Multiple Transaction Information

Year	Number of properties bought/sold by multiple transactors	Median transaction price
2006	17,212	\$57,000
2007	20,118	\$34,667
2008, Q1	4,319	\$21,495
2008, Q2	4,775	\$18,000

A transaction is included above if it was bought/sold by a party that had multiple property transactions during that year (or quarter). While the volume has remained relatively stable, the prices of this portion of the market have declined steadily. We also noticed distinct patterns in terms of participation in this market, shown below.

Market Participation Examples: Multiple Transactions

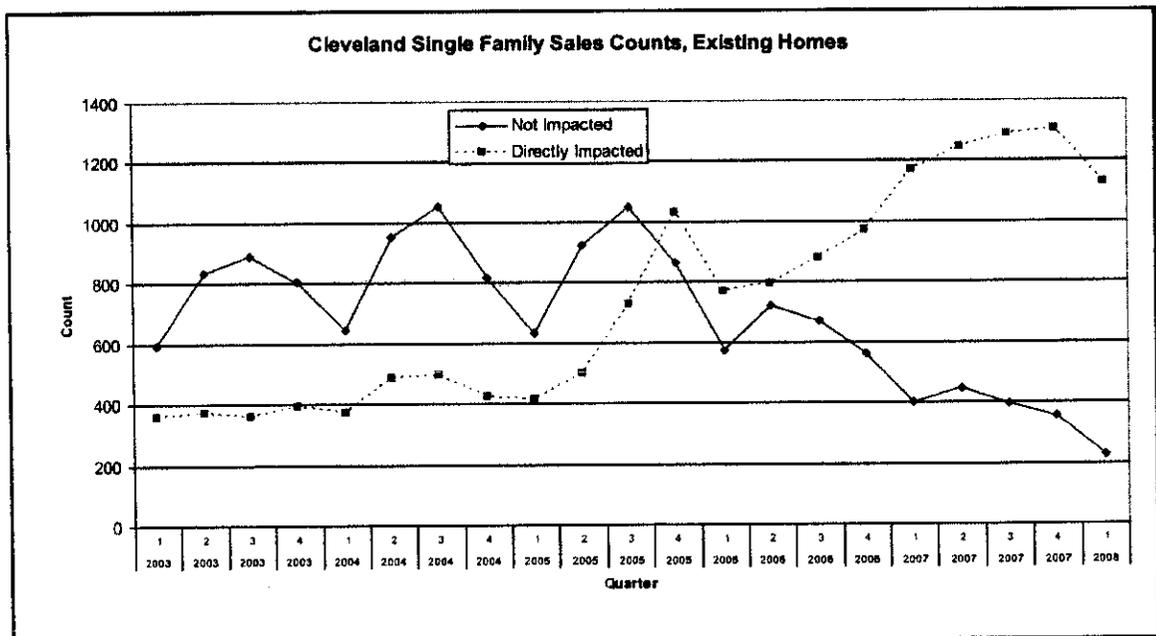
Transactor	Year	Number Purchased	Median Purchase Price	Number Sold	Median Sale Price
Rhakha, LLC	2006	25	\$15,000	19	\$92,000
US Bank	2007	377	\$33,334	207	\$10,000
BSB Investments	2008 (Q1)	17	\$250	15	\$1,000

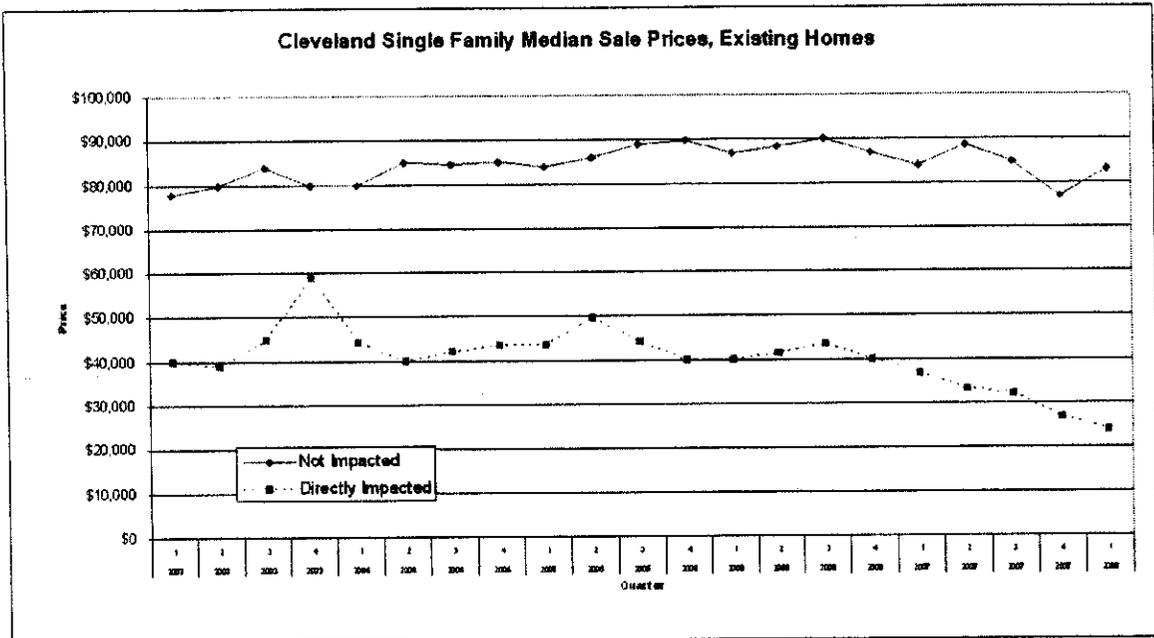
The three types of transactions represent (1) potential rehabilitation or flipping activity, (2) activity dominated by banks regaining ownership of foreclosed properties, and (3) those operating exclusively at the very bottom of the market.

Sales Prices and Volumes: We divided the single family sales market into two categories: Directly Impacted by foreclosure and Not Impacted (directly) by foreclosure. Directly impacted means one or more of the following is true:

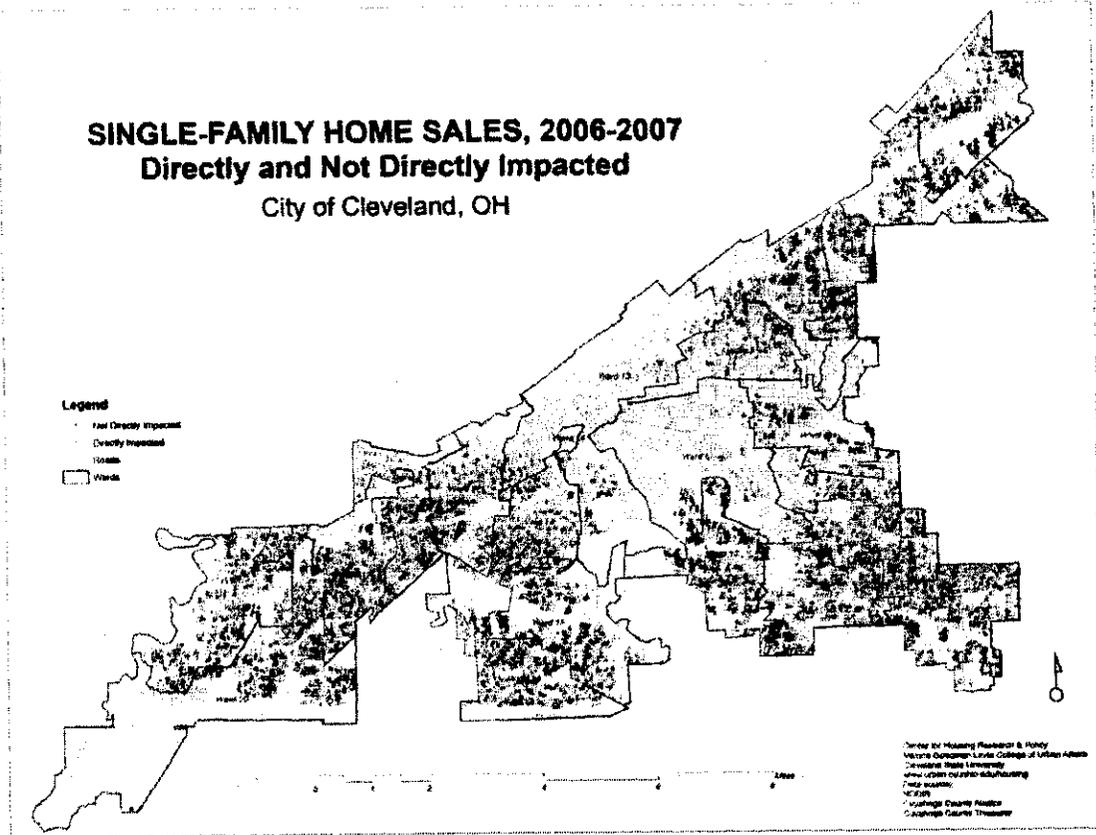
- (a) The current sale is a sheriff sale
- (b) The property was auctioned at a sheriff sale within the past two years
- (c) The property was involved in a foreclosure filing within the past two years

“Not impacted” means that none of the above conditions listed is true. We are concerned here with the *direct* impact, and graph sales price and volume by these two categories.





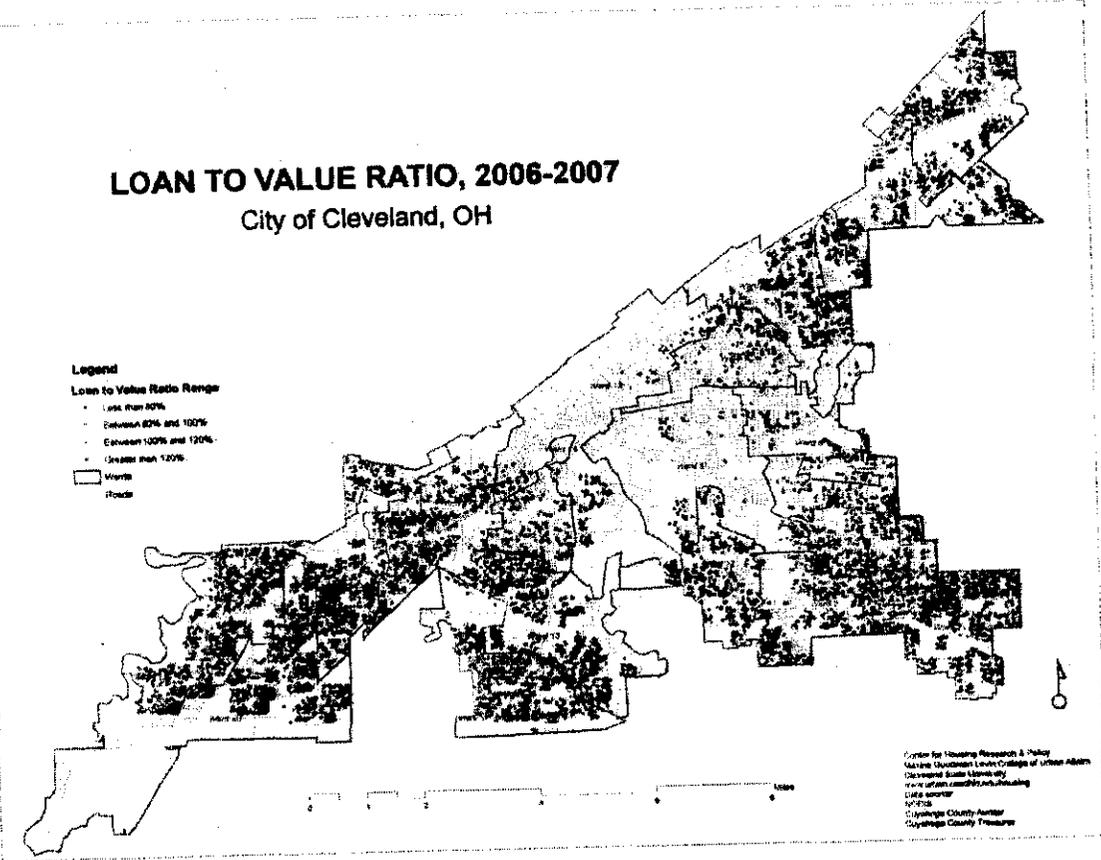
SINGLE-FAMILY HOME SALES, 2006-2007
Directly and Not Directly Impacted
 City of Cleveland, OH



LOAN TO VALUE RATIO, 2006-2007

City of Cleveland, OH

- Legend**
- Loan to Value Ratio Range**
- Less than 80%
 - Between 80% and 100%
 - Between 100% and 120%
 - Greater than 120%
- Water
- Roads



Center for Housing Research & Policy
National Government Loans College of Urban Affairs
Cleveland State University
1000 Lakeside Avenue
Cleveland, OH 44115
Cleveland County Auditor
Cleveland County Treasurer

each, but the website's speed was slow and it took time to type in each changing captcha for locating, viewing, and downloading each deed.

6. In March 2011, during my senior year, a professor at Cleveland State University consulted with me about doing some research for a lawyer. I was to test whether it would be difficult for an ordinary citizen to obtain digital copies of the Cuyahoga County's electronically-stored recorded instruments on a CD and to see how much money the county would charge as a fee to an ordinary citizen. I knew from my experience at the Federal Reserve Bank that working with a volume of recorded deeds via the county recorder's website was, as a practical matter, unworkable, so I saw the benefit of having digital copies on CD. I agreed to help. I understood that the lawyer was representing companies in a lawsuit over whether it could get copies of recorded instruments from the county recorder on CD for a reasonable fee.

7. I then went to the Cuyahoga County Recorder's office. There, I spoke with a clerk handling requests for copies of deeds and other records. I asked for a digital copy of every deed and other records that the county had recorded on four specific dates in the year 2010. I asked the clerk to put the copies on blank CDs. I said that I would pay a reasonable fee and offered to supply the blank CDs. I did not volunteer that my request was a test or had anything to do with the lawsuit.

8. The clerk told me that the recorder's office does not make copies of records on CD, and so denied my request.

9. I called the lawyer for the company suing the recorder's office – David Marburger – and told him about my experience there. Mr. Marburger told me that the county recorder already has the digital images of deeds on CD, and that it was just a matter of dubbing that CD onto a blank CD. He referred me to an online comment that he said that he'd posted on cleveland.com in response to a *Plain Dealer* story about this lawsuit. Mr. Marburger's posted comment said that the recorder's office keeps a master CD of its recorded deeds and other records.

10. Mr. Marburger also provided me with a copy of Ohio's open records law, directing me to a provision that says that a government office must provide a copy of a public record in the same format in which the government keeps that record. So if the recorder keeps copies of deeds on CD, it would have to provide copies to the public also on CD if that's what someone requests.

11. Soon afterward, in April, I returned to the recorder's office and repeated my earlier request. I spoke with the same clerk who'd rejected my request. I said that I understood the open records law to give me the right to get copies of the records on CD. Again, I did not volunteer that my request was a test or had anything to do with the lawsuit.

12. Then the clerk had me speak with a man named Larry from the recorder's computer department. I repeated my request to Larry and referred to the same part of the open records law that I'd emphasized to the clerk. Larry told me that the recorder's office does not keep any recorded instruments on CD. I told Larry that I had

read a comment on cleveland.com that said that the county recorder *does* keep copies on CD. Larry responded with a remark the gist of which was that he didn't want to get involved legally. He led me to an administration office, where I left four blank CDs along with my request in writing, my cell phone number, and my e-mail address. Someone there said that they'd get back to me about my request.

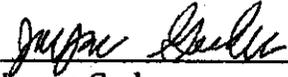
13. As more than two weeks passed, I heard nothing from the recorder's office. Then, on May 2, 2011, I received an e-mail that said it was from Larry Patterson of the recorder's office. A copy of that e-mail is marked as Exhibit 28.

14. Mr. Patterson's e-mail said: "The Cuyahoga County Recorder's Office will copy the documents you requested onto one or more cd-roms upon payment of the \$2-per-page statutory fees required under R.C. 317.32(I). As your request indicated that you will pay a 'reasonable fee' for the requested copies, please confirm that you will pay the \$2-per-page fee and we will prepare the copies you requested."

15. I replied to Mr. Patterson by asking how much his office would charge me for the CD copies at that \$2-per-page fee. A copy of my e-mail reply is marked as Exhibit 30.

16. Mr. Patterson replied by e-mail the next day. His answer: \$35,568 for 4 dubbed CDs. A copy of that e-mail is marked as Exhibit 31.

I declare under penalty of perjury that the foregoing is true and correct.

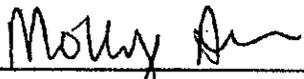


Jayson Gerbec

SWORN TO AND SUBSCRIBED to before me on this 13 day of May 2011.



MOLLY A. GILLAHAN
Notary Public, State of Ohio
My Commission Exp. Jan. 22, 2013



Notary Public