Armchair Auditing of Insolvency Processes

submitted in partial fulfilment for the degree of master of science

Tom Akkermans
11323671
master information studies
data science
faculty of science
university of amsterdam
2019-03-20

	First Supervisor	Second Supervisor
Title, Name	Dr Maarten Marx	
Affiliation	UvA, FNWI, IvI	
Email	maartenmarx@uva.nl	.

Universiteit van Amsterdam

Amsterdam Data Science

Armchair Auditing of Insolvency Processes

Tom Akkermans
University of Amsterdam
tom.akkermans@student.uva.nl

ACM Reference Format:
Tom Akkermans. 2019. Armchair Auditing of Insolvency Processes. In *Proceedings of 2 (University of Amsterdam).* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

When a company is declared bankrupt by the court, a court committee appoints an administrator to settle the bankruptcy. The administrator's task is to liquidate the company's estate and use the proceeds to settle the creditors claims. A supervisory judge ensures that the administrator is acting in the best interest of the creditors.

The supervisory function of the judge, the conflict of interest between the administrator and creditors and the appointment process of the administrator are processes which have been the subject of research[9], about which media articles have appeared [10, 11, 13] and which have led to legal proceedings

legal proceedings citation here

.

describe additional points of interest

The involved parties demand more transparency of these processes. Supervisory judges working in a reactive mode under the work pressure could benefit from data driven supervision. Information access to the general public and journalists to this processes would provide additional checks and balances to further process improvement.

The Dutch government started in 2005 publishing insolvency data[2] according to the insolvency law [3]. It provides an on-line search form [7] to retrieve a single insolvency case and provides open data web services to provide court publications and administrator reports in XML and PDF format. However, the information from a single insolvency case is limited as it does not provide aggregated and linked information. The administrator reports are unstructured and not searchable and not all interested parties can deal with the offered raw data APIs.

Instead of open data, there is a need for open analysis to enable 'armchair audits'[12] of insolvency processes. In this thesis we investigate **whether it is possible to build a complete and correct structured information system based on open and public data that is useful in that it enables such audits and search for non-technical users.**

We describe the steps in building such a system that takes in large amounts of open and publicly available data in structured and unstructured data form, extracts and enriches useful facts and makes it consumable for analysis to provide insights into the insolvency processes.

University of Amsterdam, 3,
2019. ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

2 RELATED WORK

add literature related to the research questions: on information system quality, entity linking, de-duplication, tobit model

3 METHODOLOGY

3.1 Description of Data Sources

3.1.1 *Central Insolvency Register.*

Data suppliers. The CIR [2] is operated by the Dutch government and contains company insolvency data supplied by the courts and the administrators. Courts are obliged to supply the insolvency data and free consultation thereof according to the insolvency law, article 19 [3]. CIR started the digital register on the 1st of January 2005 and retains insolvency cases until six months after the ending of the insolvency. CIR also contains other data such as personal debt restructuring (*schuldsanering*), personal insolvency and company's failure to pay (*surseance*) but this data is out of scope.

Entity records. The CIR register contains the following entities in numbers of records (as of 2019-03-21):

Table 1: number of entity records.

Entity	no. of records
Court	11
Supervisory Judge (distinct names)	580
Insolvency	51,392
Administrator (distinct names)	58,201
Publication	142,172
Report	357,803
... progress report	237,657
... financial attachment.	120,146

Publications on an insolvency case are done by the court and include the initial declaration of bankruptcy. Administrators periodically submit progress reports as well as financial attachments to the CIR.

Entity identifiers. The web service response in XML is semi structured data. It provides natural unique identifiers for Insolvency Cases, Publications and Reports so they can be easily stored in normalized SQL tables and linked. The other entities: Courts, Judges and Administrators have no identifiers but consist of free text fields for their name parts. These entities must be de-duplicated en linked to a master data record. It can be easily observed in table 1 that this is certainly needed for administrators

state estimated number of administrators

.

Entity relations. Figure 1 below shows the relationships between the entities including their cardinality. Note that some relationships are time dependent, e.g. a judge can be replaced during the lifetime of an insolvency case. Since 1-1-2019 there can be two judges appointed to one case.

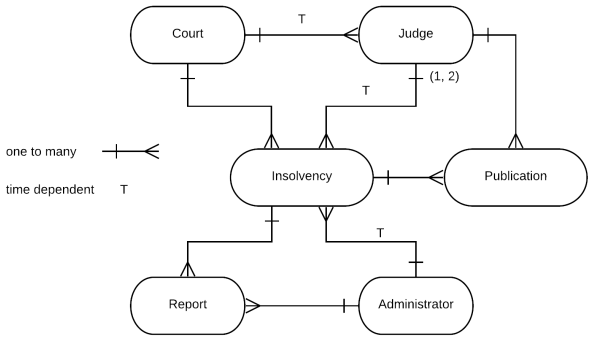

Figure 1: Insolvency entity relations.

Administrator Reports. A second web service operated by CIR provides administrator reports in PDF format. These reports hold much of the unstructured data. Recofa has published templates for both progress and financial attachment reports [6] which provide a certain structure to the contents.

3.1.2 Register of lawyers, NOvA Tableau. The NOvA tableau is the official register for lawyers and maintained by the *Nederlandse Orde van Advocaten (NOvA)*[5]. Lawyers are obliged to be registered in the tableau by the lawyer's law (*advocatenwet*, article 1 [4]). NOvA offers an on-line search form where keyword search and filters can be applied to search for a lawyer. This data source was chosen to collect the master data for Administrators.

3.1.3 Register of judges, Nevenfuncties van rechters. The Register for ancillary positions for judges is made available by *de Rechtspraak*[1]. It offers an on-line form and returns the name, current and historical occupation and ancillary positions. This data source was chosen to collect the master data for Judges.

3.2 Information System Description

Figure 2 gives an overview of the system components for sourcing, extracting, enriching and integrating the data and making the resulting structured and higher level information available to the user's analysis.

Data flows from left to right through the following components:

Data Sources. Data is sourced from three public registers:

(1) The Central Insolvency Register (*Centraal Insolventie Register or CIR*). CIR exposed both an XML and PDF file web service.
(2) The Register of lawyers (*NOvA's Tableau*).
(3) The Register of ancillary positions of judges. (*Register van nevenfuncties van rechters*)

The CIR provides the bulk of the data. The other two registers are used for the entity resolution of administrators and judges.

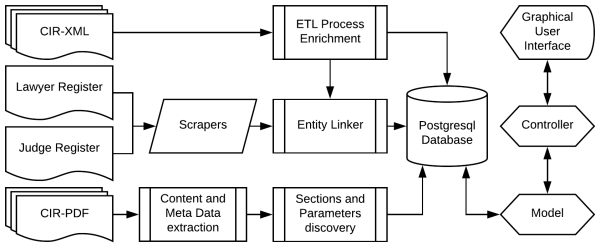

Figure 2: System overview.

ETL and Enrichment. This component loads entities with selected data fields from the CIR XML data. The data is cleaned and enriched after which it is stored in a relational database.

Entity Linker. This component is responsible for linking judges and administrators in the CIR XML data to real life entities found in the judge and lawyer registers.

PDF Processors. These components processes the CIR PDF reports to extract textual content and meta data. The text sections as defined in the progress report template and key data parameter are discovered in a subsequent process and loaded into the relational database.

Database and File Storage. Entity data is stored in a relational Postgresql database. Administrator PDF reports are stored in Amazon's S3 object storage.

Model-View-Controller (MVC). This well established pattern of subcomponents works together as a graphical interface for the user to analyse the data.

REFERENCES

[1] [n. d.]. Beroepsgegevens en nevenfuncties van rechters. https://namenlijst.rechtspraak.nl
[2] [n. d.]. Centraal Insolventieregister. https://insolventies.rechtspraak.nl
[3] [n. d.]. Insolvency Law - Faillissementswet. http://wetten.overheid.nl/BWBR0001860/2018-07-01
[4] [n. d.]. Lawyer Law - Advocatenwet. http://wetten.overheid.nl/BWBR0002093/2018-07-25
[5] [n. d.]. Nederlandse Orde van Advocated. https://www.advocatenorde.nl/over-de-nova
[6] [n. d.]. Recofa-richtlijnen. https://www.rechtspraak.nl/Voor-advocaten-en-juristen/Reglementen-procedures-en-formulieren/Civiel/Insolventierecht/Paginas/Recofa-richtlijnen.aspx
[7] [n. d.]. Zoeken in het Centraal Insolventieregister. https://insolventies.rechtspraak.nl/#!/zoeken/index
[8] 2017. *Klantwaardering Rechtspraak 2017.* Technical Report.
[9] Burak Bölük. 2011. *Is de benoeming van de curator door de rechtbank en het toezicht op de curator door de rechter-commissaris aan een verbetering toe?* Master's thesis.
[10] Dennis Meneer. 2017. Curatoren vluchten uit lege boedels. https://www.ftm.nl/artikelen/curatoren-slaan-op-de-vlucht-voor-lege-boedels Accessed on 2018-06-14.
[11] Dennis Meneer. 2018. Dit is de schadelijkste wet van Nederland. https://www.ftm.nl/artikelen/faillissementswet-schadelijkste-wet-van-nederland Accessed on 2018-06-14.
[12] Daniel E. O'Leary. 2015. Armchair Auditors: Crowdsourcing Analysis of Government Expenditures. *JOURNAL OF EMERGING TECHNOLOGIES IN ACCOUNTING* 12 (jul 2015), 71–91. https://doi.org/10.2308/jeta-51225
[13] Jan-Hein Strop. 2015. De schimmige benoeming van curatoren. https://www.ftm.nl/artikelen/de-schimmige-benoeming-van-curatoren Accessed on 2018-06-14.