

1 ARMCHAIR AUDITING OF INSOLVENCY PROCESSES
2 SUBMITTED IN PARTIAL FULFILMENT FOR THE DEGREE OF MASTER OF SCIENCE
3 TOM AKKERMANS
4 11323671
5 MASTER INFORMATION STUDIES
6 DATA SCIENCE
7 FACULTY OF SCIENCE
8 UNIVERSITY OF AMSTERDAM
9 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, 6 pages. <https://doi.org/10.1145/nnnnnnnn.nnnnnnnn>

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[2], about which media articles have appeared [10, 9, 15] and which have led to legal proceedings. The involved parties demand more transparency of these processes. Supervisory judges with significant work load could benefit from data driven supervision. Information access to the general public and journalists to this processes would provide additional checks and balances. We derive so-called Personas to represent these parties and articulate their information needs.

In 2005 the Dutch government started a digital register of insolvency data[3]. An on-line search form [16] is provided to retrieve a single insolvency case and there are web services to retrieve 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 (IS) based on open and public data that is useful in that it enables such audits and search for non-technical users.**

For this we investigate whether we can build a system that:

- constructs a complete, cleaned and fully linked entity network of insolvency cases, administrators, judges, etc.
- fully extracts the information of interest from entity data.
- fully extracts the information of interest from unstructured documents.
- provides the persona with easy to use functionality to answer its information need.

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 via a web GUI.

2 RELATED WORK

add literature related to the research questions: on entity linking, de-duplication, on survival analysis

2.1 Evaluation of information systems

Shannon and Weaver [14] state that output of an IS can be defined at different levels, at the:

- (1) **technical level**: the accuracy and efficiency of the system.
- (2) **semantic level**: the success of information conveying the intended meaning.
- (3) **effectiveness level**: the effect of information on the receiver

This implies that factors contributing to a succesful IS could also be defined at different levels. DeLone and McLean[4, 5] have gathered many factors from literature and grouped them into six distinct but interdependent categories, see figure 1 below.

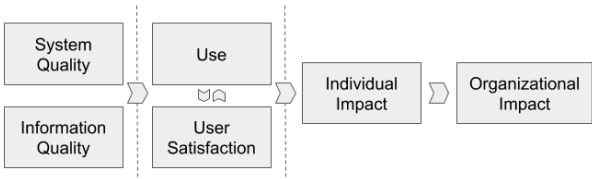


Figure 1: DeLone and McLean Information Success Model.

Each category or aspect of the I/S success measures still contains a number of variables that contribute to success:

- **System Quality**: measures the IS itself on the technical level. Example variables are data currency, completeness, and ease of use.
- **Information Quality**: measures the IS output such as reports at the semantic level. Example variables are accuracy, precision and relevance.
- **Information Use**: measures the actual use of the IS which is where the effectiveness level starts.
- **User Satisfaction**: measures satisfaction from the perspective of the user, usually on an interval scale. This category is the mostly used as a single success measure and is fairly subjective.
- **Individual Impact**: measures the information effect on the behaviour of the recipient.
- **Organizational Impact**: measures the information effect on the performance of the organization.

3 METHODOLOGY

3.1 Description of Data Sources

3.1.1 Central Insolvency Register.

Data suppliers. The CIR 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 [7]. CIR contains insolvency cases from the 1st of January 2005 and retains these until six months after they end.. 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. CIR entity data is made available by a SOAP web service returning XML responses. The XML is semi structured data and contains entities by composition which are extracted using a parser. 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 2 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.

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 [13] which provide a certain structure to the contents.

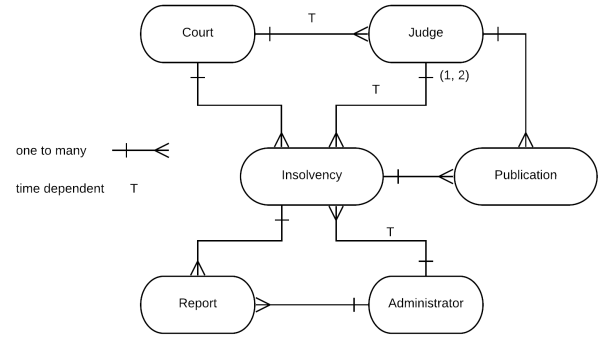


Figure 2: Insolvency entity relations.

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)* [11]. Lawyers are obliged to be registered in the tableau by the lawyer's law (*advocatenwet*, article 1 [8]). 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 3 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.

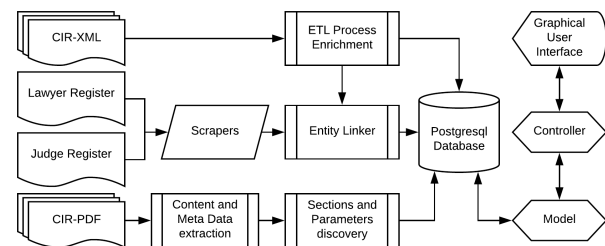


Figure 3: System overview.

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.

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.

3.3 Methods

3.3.1 Constructing the entity network. Here we describe how each entity is cleaned, standardized, de-duplicated and identified after which it can be linked to other entities to construct the entity network. For each entity a so-called master data table is created that contains the agreed upon instance or 'golden record' of the entities that can be used as a common reference point.

Courts. Courts are specified in the XML response as a free text field. A master table is created using the given names.

court name	cases
Rechtbank Amsterdam	4601
Rechtbank Den Haag	4851
Rechtbank Gelderland	5947
Rechtbank Limburg	3277
Rechtbank Midden-Nederland	5580
Rechtbank Noord-Holland	3543
Rechtbank Noord-Nederland	4556
Rechtbank Oost-Brabant	5123
Rechtbank Overijssel	3789
Rechtbank Rotterdam	5638
Rechtbank Zeeland-West-Brabant	4459
None	28

Consistency: there are no duplicates. Completeness: 28 records have missing data.

Insolvency cases. Insolvency cases are supplied with a natural identifier (*insolventienummer*), for example F.19/13/123, which consist of the insolvency type, a system number, year of insolvency and a serial number. A number of issues introduce unwanted insolvent duplicates:

- (1) data errors: padded zeros on the serial number, e.g. '0123'

- (2) redefinition of courts¹: cases administered both with the new case number format as well as the old case number format without the system number.
- (3) cases handed over to another court.

To solve for this data errors were corrected and duplicate cases were linked to the correct cases using scripts, annotating lineage and re-linking publications

Judges.

Administrators.

3.3.2 Extracting information from entity data.

3.3.3 Extracting information from unstructured reports.

REFERENCES

- [1] *Beroepsgegevens en nevenfuncties van rechters.* URL: <https://namenlijst.rechtspraak.nl>.
- [2] Burak Bölük. "Is de benoeming van de curator door de rechtbank en het toezicht op de curator door de rechter-commissaris aan een verbetering toe?" MA thesis. 2011.
- [3] *Centraal Insolventieregister.* URL: <https://insolventies.rechtspraak.nl>.
- [4] W.H. DeLone and E.R. McLean. "Information System Success: The Quest for the Dependent Variable". In: *Information Systems Research* 3.1 (1992), pp. 60–95.
- [5] W.H. DeLone and E.R. McLean. "The DeLone and McLean Model of Information System Success: a Ten-Year Update". In: *Journal of Management Information Systems* 19 (2003), pp. 9–30.
- [6] *Herziening gerechtelijke kaart.* URL: <https://www.om.nl/organisatie/herziening/>.
- [7] *Insolvency Law - Faillissementswet.* URL: <http://wetten.overheid.nl/BWBR0001860/2018-07-01>.
- [8] *Lawyer Law - Advocatenwet.* URL: <http://wetten.overheid.nl/BWBR0002093/2018-07-25>.
- [9] Dennis Meneer. *Curatoren vluchten uit lege boedels.* Accessed on 2018-06-14. 2017. URL: <https://www.ftm.nl/artikelen/curatoren-slaan-op-de-vlucht-voor-lege-boedels>.
- [10] Dennis Meneer. *Dit is de schadelijkste wet van Nederland.* Accessed on 2018-06-14. 2018. URL: <https://www.ftm.nl/artikelen/faillissementswet-schadelijkste-wet-van-nederland>.
- [11] *Nederlandse Orde van Advocaten.* URL: <https://www.advocatenorde.nl/over-de-nova>.
- [12] Daniel E. O'Leary. "Armchair Auditors: Crowdsourcing Analysis of Government Expenditures". In: *JOURNAL OF EMERGING TECHNOLOGIES IN ACCOUNTING* 12 (2015), pp. 71–91. DOI: 10.2308/jeta-51225.
- [13] *Recofa-richtlijnen.* URL: <https://www.rechtspraak.nl/Voor-advocaten-en-juristen/Reglementen-procedures-en-formulieren/Civiel/Insolventierecht/Paginas/Recofa-richtlijnen.aspx>.
- [14] C.E. Shannon and W. Weaver. *The Mathematical Theory of Communication.* University of Illinois, 1949.
- [15] Jan-Hein Strop. *De schimmige benoeming van curatoren.* Accessed on 2018-06-14. 2015. URL: <https://www.ftm.nl/artikelen/de-schimmige-benoeming-van-curatoren>.

¹in 2013 courts were consolidated (*Herziening Gerechtelijke Kaart*, see [6]).

- ²⁴³ [16] *Zoeken in het Centraal Insolventieregister*. URL: <https://insolventies.rechtspraak.nl/#!/zoeken/index>.
²⁴⁴