32nd Electric Vehicle Symposium (EVS32)
Lyon, France, May 19 - 22, 2019

EV Charging Data Management, five issues to solve

Maase, S.J.F.M.[1], Hoed, van den, R.
University of Applied Sciences Amsterdam, Weesperzijde 190, 1000 BA Amsterdam, Netherlands
s.j.f.m.maase@hva.nl

Summary

The transition towards electric mobility is expected to take off the coming years, as more EV car models access the market and charging infrastructure is being expanded. The expansion of charging infrastructure will have to accelerate to keep pace with the fast-growing need for charging. The coming years will be marked by uncertainty regarding technological developments (batteries, range), charging technologies (e.g. fast charging, inductive), growth of car sharing and autonomous driving and impact on user preferences and charging behaviour Data management is key to the EV market and public parties involved: to be able to adapt quickly to changes and to reduce risks and costs. This paper describes the five most important preconditions for effective data management that allows stakeholders to monitor the performance of their charging infrastructure and to take informed decisions on rollout strategies based on data science research results.

Keywords: charging, data acquisition, infrastructure, research, data science.

1 Introduction

The expanding market for electric vehicles and charging infrastructure asks for access to knowledge and information on charging infrastructure performance. It is important to be able to adapt to a quickly changing reality, in order to minimize business risks and social costs and to be able to assess progress with regard to climate goals. Bringing together large amounts of charging data of different sources and managing these data in a way that makes them reliable and accessible, creates not only the opportunity to assess the current performance of charging infrastructure but also provides the opportunity to develop simulation models based on data analysis. Future scenarios can be simulated and inform stakeholders on potential effects or risks of rollout policies. Data management is key to create access to this knowledge and information.

Five issues to solve in EV data management are distinguished: 1. Data collection, 2. Data quality, 3. Structuring the data, 4. Enriching the data, 5. Data Access. Solutions to these issues are described, based on the experience of managing the charging data of over eleven thousand public charging points. The current data mart contains over twelve million sessions and represents a history of more than five years of charge infrastructure rollout. This paper provides powerful insights and practical knowledge for charge point operators, mobility service providers, municipalities and all other relevant stakeholders in the field of electric mobility on how to design and optimize their data management assets. This paper provides a structured overview of the setup of a data warehouse and data mart designed to store and access charging transactions. The data warehouse and data mart enable monitoring and data access in a secure environment.

1.1 Data management

Data management includes all disciplines related to managing data as a valuable resource. This paper describes the case of data collection and data management of public electric vehicle (EV) charging infrastructure. Specifically this paper builds on charging transactions data acquired on a substantial part of the public EV charging infrastructure in the Netherlands: the four largest cities in the Netherlands (Amsterdam, Rotterdam, The Hague, Utrecht) and their regions (Provinces of Noord-Holland, Zuid-Holland, Flevoland, Utrecht). These local governments demanded their concessionaires (charge point operators (CPOs) and service providers (eMSPs)) to make available the charging data of their charging infrastructure for monitoring and research purposes. The Amsterdam University of Applied Science (AUAS) was commissioned to collect and manage the charging data. A reliable data warehouse (DWH) and data mart (DM) was developed and kept up-to-date on a monthly basis since 2014. At the time of writing (March 2019) the DM supports data analysis of twelve million valid EV charging sessions executed at over eleven thousand public charging points in the western part of the Netherlands.

1.2 Charging data

The charging data comprise of periodically (once a month) supplied transaction files consisting of Chargepoint Detail Records (CDRs) and Meter Values (MVs) of each public charging location part of the concessions in the cities and regions mentioned above. CDRs and MVs are delivered in the following data formats: csv, xls, xlsx, json or xml, depending on the concessionaire's (be it the CPO or eMSP) preference.

Table 1.1: Charge point Detail Record (CDR): column names and content

CDR_ID	Start_datetime	End_datetime	Duration	Volume	Charge_Point_Address	Charge_Point_ZIP	Charge_Point_City	Charge_Point_Country	Charge_Point_Type	Product_Type	Tariff_Type	Authentication_ID	Contract_ID	Meter_ID	OBIS_Code	Charge_Point_ID	Service_Provider_ID	Infra_Provider_ID
1601312350438xxxxxx	2016-01-31T21:20:50+01:00	2016-01-31T23:50:43+01:00	2:29:53	4,31	Nicolaas Witsenkade 25	1017ZS	Amsterdam	NLD	3	0	0	04EC35C2xxxxxx	NL-NUO-xxxxxx-X			EVB-P1541099_2	Nuon Consumer	Nuon
etc.																		

```
"cdr_id","interval_stop","meterwaarde"
"4554480","2019-01-31 22:31:13","1040"
"4554480","2019-01-31 22:33:05","2400"
"4554480","2019-01-31 22:35:00","3810"
"4554480","2019-01-31 22:36:51","5210"
```

Figure 1.1: example of metervalues delivered as .csv

```
<MeterValue>
    <ChargePoint>DB0339</ChargePoint>
    <CollectedValue>16</CollectedValue>
    <Connector>2</Connector>
    <InternationalName>NL-EVN-DB0339-
2</InternationalName>
    <TimeStamp>2019-03-
11T23:00:00Z</TimeStamp>
    <TransactionID>3336256</TransactionID>
  </MeterValue>
```

Figure 1.2: example of metervalues delivered as .xml

2 Five data issues

2.1 Data collection: different data file formats and API calls

In the case of EV charging data multiple protocols are in place in the Netherlands to exchange information on authorization identification, charge point location, billing and charging infrastructure control. Figure 2.1 visualizes existing protocols in relation to the various stakeholders in the EV charging market. This figure is an adjusted version of the figure shown and described in the “EV related protocol study 1.1” executed by ElaadNL in 2017 [1]. Historically the charging data gathered for data analysis by the Amsterdam University of Applied Science (AUAS) were not delivered applying any exchange protocol. File formats and API calls differ for each provider of charging data. Section 3.1 describes how the issue of various file formats and API calls has been solved. In Q1 2019 testing of the OCPI protocol for data-exchange is being executed. The OCPI protocol standardizes file formats and API calls. This should lead to a highly automized data collection process including more frequent updates of the data.

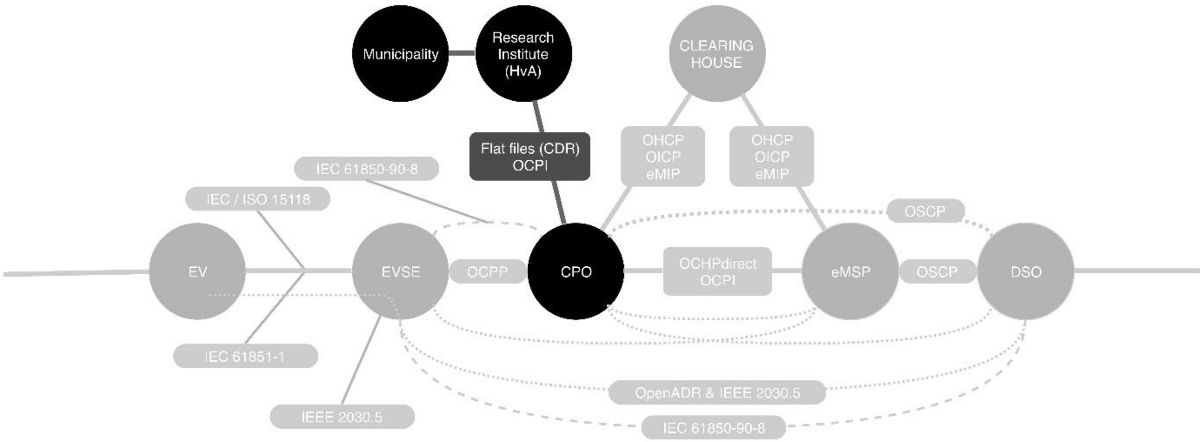

Figure 2.1: EV Data exchange protocols in the Netherlands.

2.2 Data quality: empty cells, errors and inconsistencies

The second issue is the issue of data quality. Missing, incorrect or inconsistent data can lead to false research results or wrong interpretations of EV charging infrastructure performance. Table 2.1 lists the type of errors and inconsistencies seen in charging data. Section 3.2 elaborates on how these errors and inconsistencies are dealt with.

Table 2.1: possible errors and inconsistencies in EV charging data.

missing data	Any cell can be empty
errors	Location errors: Region <> District <> subdistrict <> subsubdistrict <> Location Address kWh > 100 kWh or kWh < 0 Invalid date eg. date in the future StartConnectionDateTime > EndConnectionDateTime Negative Connection Time StartConnectionDateTime = EndConnectionDateTime
inconsistencies	Double rows within batch Late arriving double rows Broken session StartConnectionDateTime < previous EndConnectionDateTime chargepointID <> Region <> District <> subdistrict <> subsubdistrict <> Location Address

2.3 Structuring the data for research purposes

EV charging data back in the early years of collecting by the AUAS, were organized in a so-called flat file database. Only charging sessions were reported. These sessions were analyzed as flat files in Microsoft Excel for the municipality of Amsterdam only. Soon the expansion of both EV charging locations and number of partners (the four largest cities in the Netherlands and the metropolitan regions) led to an amount of data MS Excel could not process properly. Research questions transformed from descriptive analysis to more complex statistical analysis and the construction of simulation models. For these research purposes it is required to be able to combine and compare data in new ways. Innovative forms of information processing that enable enhanced insight, decision making, and process automation have been developed over the past five years. Structuring and restructuring the data has been a continuous process. Data managers working closely together with researchers to create the desired flexibility to analyze the data. Section 3.3 describes the structure of the data warehouse and data mart providing the opportunity to combine data in a flexible way.

2.4 Transforming research contexts ask for contextual data

The field of electric mobility has been developing steadily over the past decade. As earlier policies focused at improving air quality by stimulating electric mobility and stimulating the purchase of EVs, policies are increasingly focused on stimulating the roll out of charging infrastructure, while managing impacts on energy grids and matching EV charging demands with renewable energy generation through smart charging strategies. Another policy trend is a more diversified approach focusing on specific user groups like taxis, freight, and car sharing programs. These shifts lead to research questions that require more information than charging data per se: other data such as lay out of electricity grids, renewable energy production and travel patterns of particular target groups for example. Section 3.4 describes relevant contextual datasets integrated in the data warehouse and data mart of charging data serving these changing research contexts.

2.5 Data Access: securing sensitive data

Charging data contain business and privacy sensitive information. *Business sensitive information* may include the volume of charging sessions (kWh charged per session) providing insight in the business case of a charging location or the charging infrastructure maintained by one specific CPO. *Privacy sensitive information* may include authorization indicators like charging card RFIDs. Chances are that if combined with other -non- EV charging datasets identity of EV users might be revealed without their consent. The municipalities and region's goal to collect charging data was to support the public interest of clean air and energy transition by monitoring the rollout process and charging infrastructure usage. Access to the data has to be secured properly to protect partners' and individuals' interests. Data management, security and privacy of the involved has to be compliant with European Legislation. The concept of privacy by design is applied. Section 3.5 expands on how sensitive data are secured and shared at three different levels taking into account the public, business and privacy interests.

3 Solving data management issues

3.1 Data collection

CPOs, MSPs and other stakeholders involved in the primary process of exploiting EV charging infrastructure exchange charging data. Data management is primarily targeted at the core business process: authorization, control and billing of the EV charging services delivered. Protocols have been developed and implemented to support this primary business process (Figure 2.1 in section 2.1). The goal of data collection and -management as described in this paper is to create the possibility to analyze charging infrastructure performance and to develop models to forecast and simulate future scenarios. The following problems arise when gathering charging data for research purposes: [i] sources and channels through which the charging data are delivered vary and [ii] data formats vary (Figure 3.1).

As soon as the borders of MS Excel had been reached with regard to the amount of data to process, velocity of processing and flexibility in data analysis, AUAS started to build a data warehouse structure which will be described in section 3.3. But before being able to load the data into a data warehouse, the problem of collecting from various sources and channels and various data formats had to be solved.

Because charging data have a transactional character MS SQL Server has been chosen to store and manage the data. The SQL Server Integration Services platform has been used to develop so called SSIS packages for extracting, transforming (section 3.2) and loading the charging data into the data warehouse (section 3.3). SQL Server Integration Services is a platform providing tools and built in tasks and transformations to extract and transform data from a wide variety of sources such as XML data files, flat files, and relational data sources, and then load the data into one or more destinations. Figure 3.1 shows the various sources, channels and file formats AUAS manages.

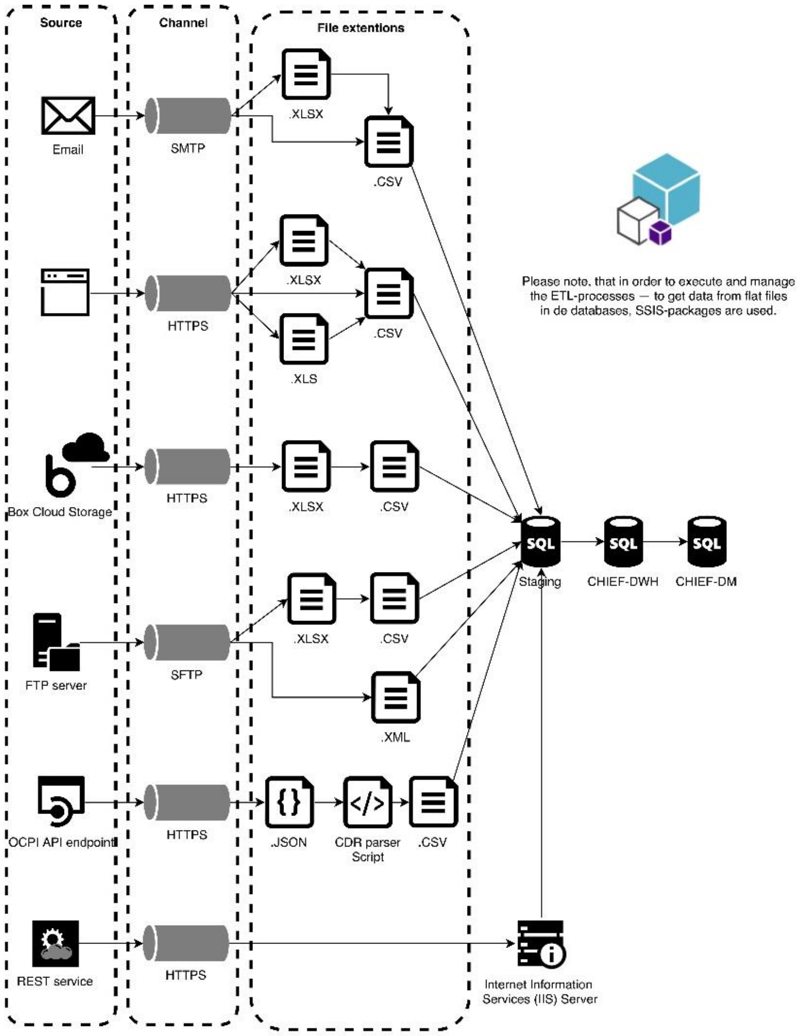

Figure 3.1: sources, channels and file extensions

The fifth row of sources and channels in figure 3.1 shows OCPI as a source of charging data collection. Recently AUAS together with CPOs started testing data collection using the OCPI protocol. Primarily the OCPI protocol has been developed as a standard protocol to exchange charging data between CPOs and eMSPs (electric Mobility Service Providers). In the case of collecting data for research purposes, implementing the OCPI protocol will change the batch wise delivery of charging data in all kinds of different file formats into a more continuous stream of data with a more uniform format. One of the hurdles to overcome to be able to receive the data through OCPI had to do with the protocol being developed for exchange of data, not for receiving data only. The OCPI protocol originally is designed with a two-phase handshake, authorizing both parties to exchange data bidirectionally. Data collection for research purposes doesn't require this two-phase handshake because the research institute only has to receive data. Bidirectional exchange is not necessary and can therefore be eliminated. Implementing the OCPI protocol for data collection for all CPOs and eMSPs provides the opportunity to extensively automate the data collection process.

3.2 Data quality

Missing, incorrect or inconsistent data can lead to false research results or wrong interpretations of EV charging infrastructure performance. Table 2.1 in section 2.2 lists the type of errors and inconsistencies seen in charging data. These errors and inconsistencies may have been caused by human entry errors or by corruption in transmission or storage. Although implementing the exchange protocol OCPI automates the collection process, data quality is not being improved per se. The current version of the OCPI protocol does not prescribe the format of content per entry.

Data cleansing or cleaning is the process of detecting and correcting (or removing) corrupt or inaccurate records from a record set, table, or database and refers to identifying incomplete, incorrect, inaccurate or irrelevant parts of the data and then replacing, modifying, or deleting the dirty or coarse data [2]. Data cleansing in this case is performed as batch processing through SSIS packages and stored procedures developed specifically to clean transaction CDRs. Errors and inconsistencies are either marked invalid or corrected based on known errors or contextual knowledge (correcting known errors). Table 3.1 lists the result of cleansing for each type of error or inconsistency. The cleansing code developed can be provided by the author upon request.

Table 3.1: cleansing result per error / inconsistency in EV charging data.

	Type of error or inconsistency	**Cleansing result**
missing data	Any cell can be empty	No action
errors	Location errors: Region <> District <> subdistrict <> subsubdistrict <> Location Address	Manual fixes based on location history
	kWh > 100 kWh or kWh < 0 kWh	Mark session invalid
	Invalid date eg. date in the future	Mark session invalid
	StartConnectionDateTime > EndConnectionDateTime	Switch of StartConnectionDateTime and EndConnectionDateTime after check with given Connection Time
	Negative Connection Time	Fix after calculation of connection time based on start and end.
	StartConnectionDateTime = EndConnectionDateTime	Check volume kWh and given connection time; if valid than calculate and repair endconnectiondatetime
inconsistencies	Double rows within batch	Mark session invalid
	Late arriving double rows	Mark session invalid
	Broken session (same RFID repetitively connected within portion of an hour)	Mark session invalid, if fastcharger=Y: session = valid. Under construction: chaining broken sessions at non-fastchargers.
	StartConnectionDateTime < previous EndConnectionDateTime	Mark second session invalid
	chargepointID <> Region <> District <> SubDistrict <> SubSubDistrict <> Location Address	Manual fixes

3.3 Structuring the data

As mentioned in section 3.1 EV charging data have a transactional character: charging sessions relate to charging locations and charging cards etc. The charging data and other relevant datasets (e.g. location data) have a transactional character. MS SQL Server was selected as the software package in which to store and retrieve the charging data as it is a relational database management system providing high capacity and performance. Figure 3.2 gives an overview of the data table structure and the relations between the data as seen by the data warehouse development team. Section 3.5 elaborates on data access and data security. Since recent research questions demand a more agile approach a transition to a data lake is now being prepared. The AZURE cloud is being considered for easier maintenance of scalability, security, availability and continuity, but not yet developed.

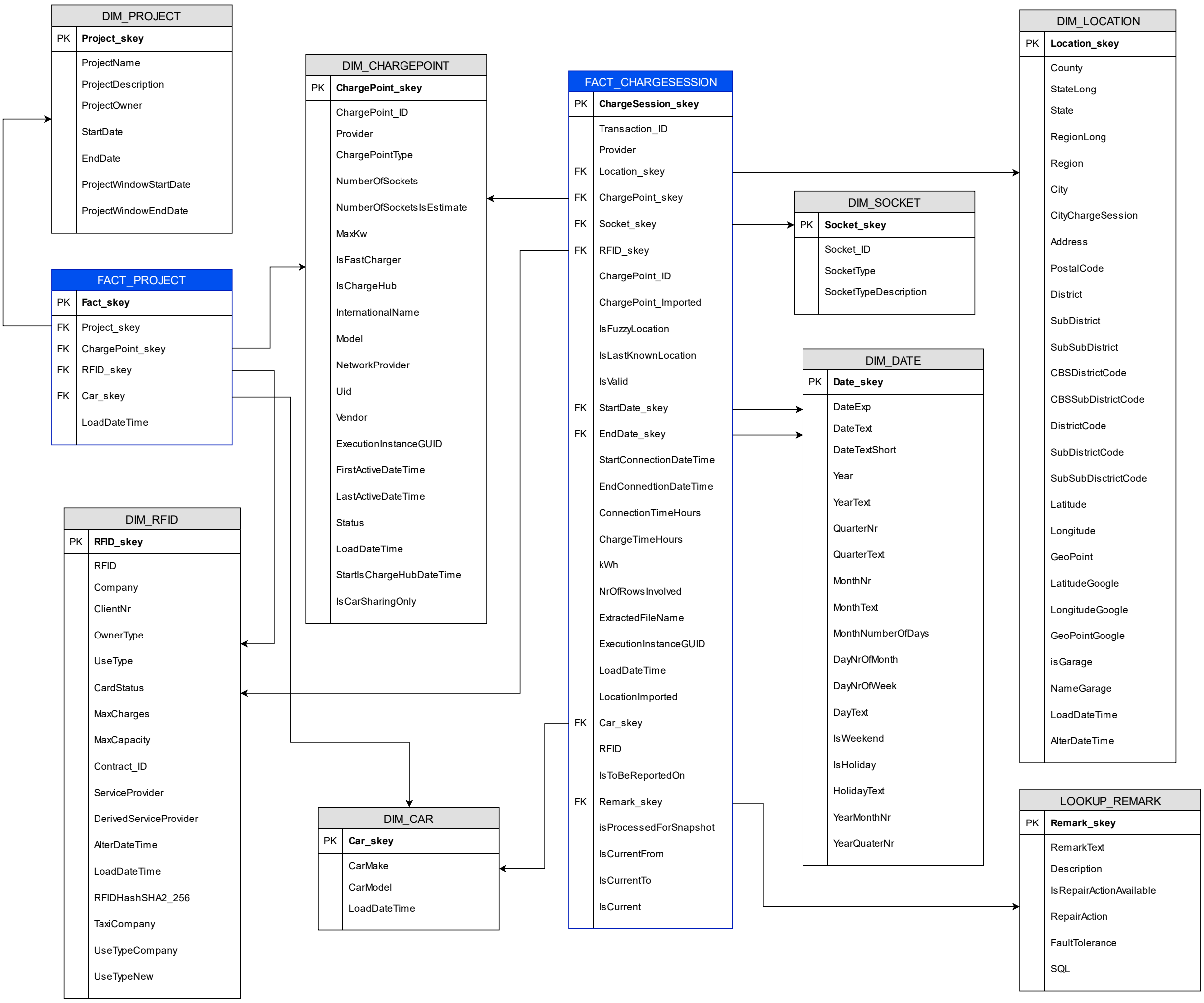

Figure 3.2: Data warehouse structure: data being organized in a relational database as seen by the developers.

3.4 Enriching the data

Charging data are an important source for evaluation of charging infrastructure performance and short-term adaptation of the rollout process. Nevertheless for [i] scaling up the roll out process and simulating future scenarios, [ii] evaluating the impact of charging infrastructure on the energy transition and [iii] policy focus on specific user groups, contextual data become increasingly relevant. As the sheer size of charging infrastructure, the market for EVs and effect on the grid grows, it becomes highly important to be able to scale up and steer the rollout process integrating knowledge and information based on contextual data.

Monitoring of charging infrastructure is generally confined to the most important key performance indicators like number of sessions, volume charged and occupancy. In order to be able to build simulation models, to be able to define specific user behavior, to calculate the impact of EV charging on the electricity grid and to research the potential of smart charging, contextual data is needed.

Contextual data in this case refers to contextual information regarding [i] location of the charging infrastructure, [ii] technological specifications of the charging infrastructure and [iii] user labels (table 3.2). Subsections 3.4.1 and 3.4.2 demonstrate the application and value of contextual data.

Table 3.2: contextual data sets

	contextual data
location	Charging location geo-coordinates District, subdistrict and sub-subdistrict markers per charging location
technological specs	Fastcharger Y/N Number of charging points available at charging location Maximum available charging power at charging location
specific user groups	Usergroup labels (eg. taxi, car sharing program) Taxistand entry and exit point registration
Energy transition	Weather data Renewable energy production data Smart charging experiment indicators per charging location Energy pricing data (APX) Energy grid data

3.4.1 Promoting clean taxis in Amsterdam

The city of Amsterdam wants to have an emission free taxi sector by 2025. In order to reach that goal, the city has taken a number of measures which favour clean taxis above conventional taxis. In 2015, as part of this policy, the city of Amsterdam decided to introduce a privilege measure for clean taxis at the Amsterdam Central Station taxi stand. In 2017, only clean taxis were allowed at the Leidseplein taxi stand, one of the busiest taxi stands at social nights in the city. In addition to the basic set of charging data, also entry and exit point data of taxistands was needed to be able to monitor and evaluate implemented policies. Once a charging card had been executed a charging session at the central station taxistand, this RFID was automatically labeled “taxi”. By doing so, general rules for charging behavior of this specific usergroup can be derived from the data and implemented in simulation models.

3.4.2 Simulating vulnerability: application of walking distance

Another example of contextual data enriching the charging data set is generated data about walking distances between EV charging stations. The walking distance script renews the walking distance data set every month including all new charging infrastructure recently installed. This contextual dataset has been applied to the so-called vulnerability script. By using walking distance as a decisive argument for EV drivers to select a charging location in case the desired location is occupied or out of order, the vulnerability of the EV charging infrastructure is calculated. This simulation report informs municipalities and CPOs where to expand charging infrastructure or which assets to pay extra attention to regarding maintenance.

3.5 Data Access

EV charging data contain business and privacy sensitive information as described in section 2.5. Data security is therefore highly important. AUAS together with the data owners – the municipalities and metropolitan region Amsterdam – developed a data sharing protocol. Three levels of data sharing have been defined: [i] open access to aggregated performance indicators at the level of municipalities and regions, [ii] access to reports and evaluation tools at individual charging station level, [iii] access to the charging data at session level (see figure 3.3: data access at level 1, 2 or 3).

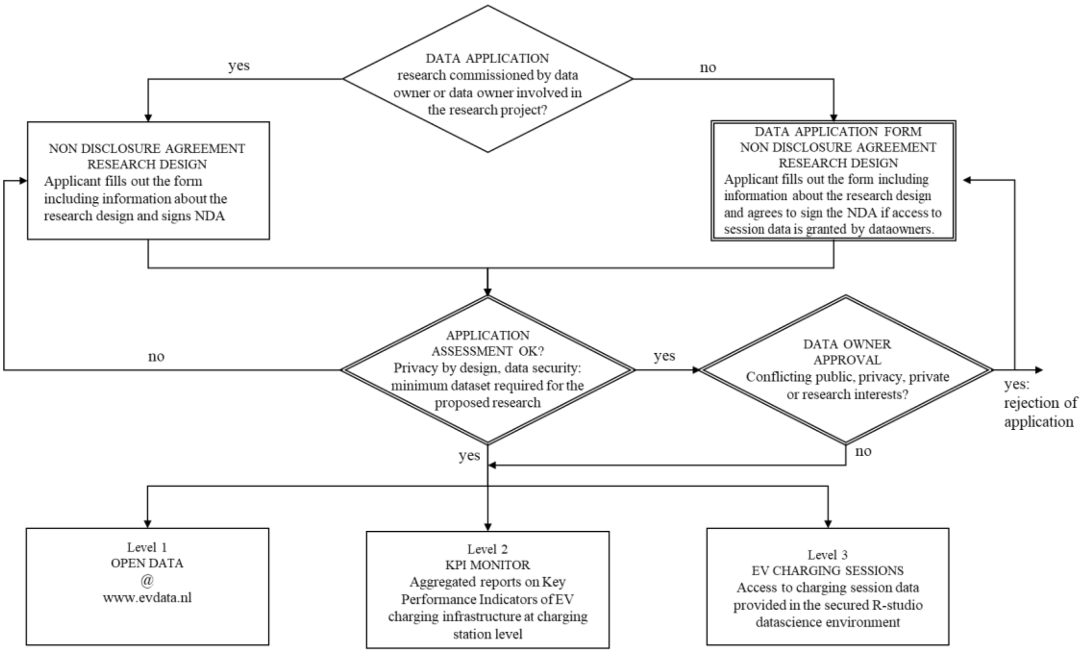

Figure 3.3: baseline data sharing protocol

3.5.1 Level 1 data access

The most aggregated level of charging data can be accessed publicly on the web [3].

3.5.2 Level 2 data access

Monitoring tools [4] showing key performance indicators at the level of individual charging stations can be accessed only by password secured personal accounts for employees of municipalities. Accounts only get access to the information about those charging stations they are allowed to see as determined by the data owners. An R-package was developed to specify the dataset on which the KPIs shown are based.

3.5.3 Level 3 data access

AUAS data scientist and developers working on projects commissioned or approved by the data owners, get access to relevant subsets of the charging data and contextual data depending on the research goals and research questions to be answered by data analysis. Data scientists and third parties granted access to the charging data at session level enter by using a three-level authentication. [i] VPN connection, [ii] R-server entry [iii] connection to CHIEF-DM data mart. An R-package was developed to enable access to specified datasets and reports (as mentioned in section 3.5.2). This specification involves access at location level, accessed timeframe, CPO level, project level and whether a data scientist can see RFID hashes. After three level authentication data scientists get access to an R-studio environment for data-analysis as shown in figure 3.4. Computing power is provided giving access to a computational server.

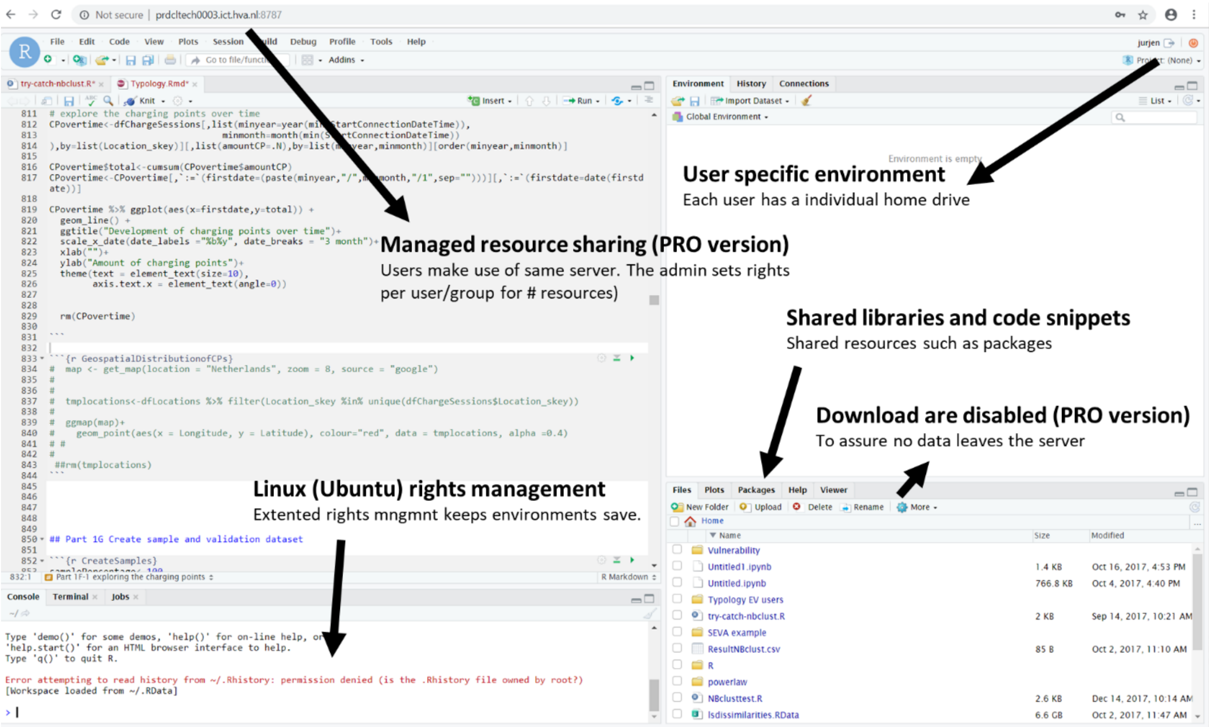

Figure 3.4: R-studio data analysis environment

Each request for access to data, if not public, is processed applying the data sharing protocol shown in figure 3.3. External parties can also request access to the charging data by filling out a data application form. The data owners, be it the municipalities and the metropolitan region decide whether an external party gets access. A data application by third parties is assessed against the background of the public, private and research interest and can be honored at all three levels of data access as mentioned above. A fourth level of sharing aggregated data which is not shown in figure 3.3 is the production of aggregated data reports for the third party on request. In this case also approval of the data owners is necessary. These reports on request will also be published publicly.

At all levels of data sharing, the data owners decide who is allowed which level of access to which subset of stored charging data.

4 Conclusions and future challenges

The Dutch case of EV charging data management proves that data management of charging data of public charging infrastructure at a national scale is feasible. The data warehouse developed and maintained by the AUAS provides secure and controlled access to EV charging data at different levels of data aggregation and secured access. Access is granted by the data owners. Data analysis at charging session level can be executed in an R-studio environment connected to a powerful computational server. The importance of bringing together both charging data and contextual data is illustrated by research outcomes, referred to in this paper. Five issues with regard to EV charging data management have been revealed and their solutions explained. Table 4.1 summarizes these issues and their solutions.

Table 4.1: EV charging data management 5 issues and their solutions

	Issue	Solution
1. Data collection	Different sources and file formats	SSIS packages for ETL
2. Data quality	Errors and inconsistencies	SSIS packages for ETL, manual corrections
3. Structuring data	Research flexibility	MS SQL server
4. Enriching data	Shifting research contexts	Contextual data stored and available for datascience
5. Data access	Data security, privacy and conflicting interests	Data sharing protocol including three levels of data access.

The availability of this unique set of historical and up-to-date EV charging data of metropolitan and regional areas in the Netherlands provides a strong advantage in monitoring and steering the renewable energy transition and the adoption of electric mobility.

Future challenges with regard to charging data management entail:

- Processing more data, real time.
- Standardizing and implementing new protocols for data exchange: the implementation of new protocol does not guarantee data quality improvement in itself. Data warehouse structures will have to be adapted to those new protocols. Data cleansing methods will have to be adapted accordingly.
- Data collection is becoming easier through protocols like OCPI and thus creating the opportunity for anyone granted access to collect and store charging data for research purposes at various locations. Application of different data management processes and data structures can lead to deviations in research outcomes based on the "same" data collected, but diversely managed. We would favor the implementation of one national charging data warehouse or charging data lake managing EV charging data and controlling data security, cleansing and access. Standardizing the data cleansing process is highly important to create unambiguous research results.
- Previously municipalities played a central role in the rollout process of charging infrastructure. A shift now is taking place from steering the rollout process actively by the local government towards controlling the performance of concessionaires. Providing access to specifically aggregated dataset to feed business intelligence tools built by the municipalities themselves more and more is requested.

Acknowledgments

The data management process has been developed in close cooperation with the municipalities of Amsterdam, Rotterdam, the Hague, Utrecht, and the Metropolitan Region Amsterdam Electric. We would like to thank all practitioners for contributing their knowledge and experience to our case study and the development of the assessment platform. The development could not have been executed without the municipalities of Amsterdam, Rotterdam, The Hague, Utrecht, and the Metropolitan Region Amsterdam Electric providing the charging data and funding. This research and development has also been funded by the Dutch "Nationaal Regieorgaan Praktijkgericht Onderzoek SIA" grant number RAAK-PRO 2014-01-121.List acknowledgments here, if appropriate.

References

[1] *ElaadNL*, https://www.elaad.nl/downloads/downloads/ accessed on 10-3-2019 elaadnl-ev-related-protocol-study-v11.pdf

[2] Wu, S., *A review on coarse warranty data and analysis,* Reliability Engineering and System, (2013)114: 1–11, doi: 10.1016/j.ress.2012.12.021

[3] *EVdata*, https://www.evdata.nl, accessed on 10-3-2019

[4] S.J.F.M. Maase et. Al., *Performance of Electric Vehicle Charging Infrastructure: Development of an Assessment Platform Based on Charging Data,* World Electr. Veh. J. (2018), 9(2), 25; https://doi.org/10.3390/wevj9020025

Authors

Simone Maase is project manager and research associate at the Amsterdam University of Applied Sciences (AUAS), part of the Urban Technology research program. Research topics include grass roots initiatives for a sustainable society and the transition towards electric mobility. Since 2014 she manages the "monitoring public EV infrastructure" and the "IDO-laad" research projects at the AUAS.

Robert van den Hoed is Professor Energy and Innovation at the Amsterdam University of Applied Sciences (AUAS), part of the Urban Technology research program. Research topics include electric mobility, optimization of charging infrastructures and smart grids. Van den Hoed is board member of Dutch-INCERT, steering group member of Amsterdam Smart City and National Knowledge platform Charge Infrastruture (NKL), and member of the scientific review committee of the European Electric Vehicle Congress