Artificial Intelligence models for the sewage water system in the province Limburg, The Netherlands

Pre-Commercial Procurement (PCP)

Market consultation results report

Waterschapsbedrijf Limburg ("Contracting entity")
Postbus 1315
6040 KH Roermond
Maria Theresialaan 99
6043 CX Roermond

Disclaimer & copyright

In conducting the Market Consultation, WBL benefited from assistance offered under the EAFIP-initiative (www.eafip.eu), which is supported by the European Commission Directorate-General Communications Networks, Content & Technology (DG CONNECT). Economic operators are being informed that any information regarding the setup and execution of both the procurement process and the execution of any contract/framework agreement as a result of the procurement process as well as public summaries of the results of the PCP project, including information about key R&D results attained and lessons learnt by the procurers during the PCP, can be shared after consultation with the respective R&D provider by WBL with(in) the context of the EAFIP-initiative and consequently can be analyzed, (re-)used and published by the EAFIP-initiative. Details should not be disclosed that would hinder application of the law, would be contrary to the public interest, would harm the legitimate business interests of the R&D providers involved in the PCP or could distort fair competition between the participating R&D providers or others on the market.

CONTENT

1. Introduction

The OMC aimed to inform market operators regarding the upcoming procurement of research and development (R&D) services for the development of artificial intelligence models for the sewage system in the province Limburg.

The OMC also aimed to understand the operators' capabilities to satisfy the procurers' needs and to obtain market parties' input on the viability of the procurement plans and conditions as described in this document and annexes.

The timetable for the OMC was set as follows:

12 September 2019	Publication of pre-announcement and making available market consultation documents & list of questions (through the EU Survey).
20 September 2019	Possible publication of updated version of the EUSurvey.
27 September 2019, not later than 15:00 CET	Registration for the meeting in Roermond
4 October 2019, 14:00 – 16:00 CET	Physical meeting for explanation and possible questions. Date: 4th of October 2019 Time: 14:00 – 16:00 CET Location: Congrescentrum Het Forum, Elmpterweg 50, 6042 KL Roermond, The Netherlands Info: http://forumroermond.nl/
12 October 2019, not later than 17:00 CET	Deadline to receive answers to the questions in the EU Survey
31 Oktober 2019	Publication of market consultation report.
1st November 2019	Closure of market consultation.

The OMC started on the date of its publication in the EU's Official Journal (TED) with closure on 1st of November 2019. The market consultation was organised in the form of a meeting and an online questionnaire. The meeting will took place on the 4th of October 2019, in Roermond, the Netherlands.

Market parties were also requested to fill out the questionnaire available at https://ec.europa.eu/eusurvey/runner/WBLBigDataPCP2019 The deadline to fill out the questionnaire was 12th of October 2019. The responses to the questionnaire could not contain any confidential information. The intention of the questionnaire was to explore the market 'as-is', therefore there could not be no wrong or right answers. The answers provided will be used as input for our procurement strategy and conditions.

After processing the questions and responses of all suppliers, this document is a means to process and communicate the results. In this context we treat all information provided by operators as commercially sensitive and we will therefore not communicate any supplier specific details. Only the general findings are summarised and communicated in this report. This anonymised report (excluding the confidential information) will be published on 1st of November 2019 via Tenderned and TED.

All communication (before, during and after the OMC) was carried out in English. However, for the subsequent procurement, the interface of a final solution should be made available in Dutch language. In connection with this requirement, several documents that were made available as annexes were in Dutch language. Participation in the OMC is not a condition for submitting a tender in the subsequent procurement. Also, the provided input in the OMC will not be used to evaluate future proposals.

By carrying out the market consultation, the procurers do not commit to subsequently deploying a procurement procedure. Moreover, in case this OMC will be followed by a procurement procedure, the public procurers reserve the right to change any elements that define the desired solution. No rights can be derived from any statements made by the procurers during the OMC.

The following annexes were provided:

- ICT-Group N.V. - Adviesrapport innovatie bedrijfsvoering afvalwatertransportstelsel
- ICT-Group N.V. - Rapportage workshops MOMS
- Witteveen+Bos - Onderzoek slimme besturing transportsysteem
- Witteveen+Bos – Set prestatie indicatoren transportsysteem

- RHDHV – Basisdocument Aquasuite FLOW pilot Venray
- RHDHV/WBL – Aquasuite Flow op ATS Venray
- Brightcape – Big Data Innovation Vision Waterschapsbdrijf Limburg
- Geographical map transport system Venray

The questionnaire (EU Survey) has been explicitly qualified as an annex to the market consultation document and is not a standalone document, but is part of a set of documents. For the sake of readability, the choice has been made to have the questions asked in the Market Consultation documents within the framework of the EU Survey (https://ec.europa.eu/eusurvey/runner/WBLBigDataPCP2019).

The summary of the feedback received to the questions of the abovementioned EU Survey questionnaire and the extract of the prior art analysis are presented in the following sections of this report.

2. Results of the EU Survey questionnaire

The following summary compiles the feedback received to the EU Survey questionnaire consisting of 17 questions (listed below). The results are presented in an anonymized manner as prescribed in the Market Consultation document. The information covers in 17 topics the envisaged planning of the PCP, the existence of patents and standards, and the relation of the Big Data PCP with the procurement of process automation and standardisation of PA systems, network and information security, and improvement of work processes consisting of three lots[1]:

- Lot 1: Process Automation;
- Lot 2: Security;
- Lot 3: Big data.

Questions:

1	Do you have any suggestions regarding the scope of the envisaged PCP that would lower the budget for the delivery of the use of Big data solutions in the water chain?
2	What would you consider an appropriate budget for delivering the services detailed in the market consultation document and related annexes. Please elaborate.
3	Do you have knowledge of any suitable technology for Big data solutions/analysis (e.g. prediction models) in the field of water management? Please elaborate.
4	Do you have knowledge of any suitable methodology/approach to estimate the impact of Big data solutions in the water chain? Please elaborate.
5	Can you suggest a classification for potential technologies (e.g. value stream design, etc.)?
6	Can you describe your approach to developing and testing technologies for Big data solutions in the water chain?
7	Do you foresee any barriers to use of existent solutions in different settings? If yes, please elaborate.
8	Do you foresee any barriers to implement Big data solutions in the water chain automation, maintenance and security? If yes, please elaborate.
9	Can you offer relevant functionalities that have not been described in the Market Consultation Document ? If yes, please elaborate.
10	Are there any inadvertencies in the requirements as described in the Market Consultation document ? If yes, please elaborate.
11	Can you provide any other recommendations regarding the Big data PCP for the development of innovative solutions for the water chain automation, maintenance and security?

[1] For the detailed description refer to the Market Consultation document and its annexes

12	Can you indicate whether the listed patents and standards in the prior art analysis are relevant?
13	Can you name additional relevant standards and patents? Are you aware of other relevant rights or trade secrets? If so, please provide reference to the revelant patent registration and details, as well as a general description of any relevant rights or trade secrets.
14	Do you own any relevant IPR to the objectives of Waterschapsbedrijf Limburg?
15	Are you aware of any patents that may constitute a barrier to your delivering a solution in the envisaged PCP procurement?
16	Can you suggest a more suitable list of developments in the field?
17	Do you commercialize Big data solutions for the water chain automation, maintenance and security? If yes, please elaborate.

2.1. Suggestions on the scope of the PCP to lower the budget

Diverse suggestions have been provided. Some point out that existing software products are unlikely to bring maximum benefit to WBL because they have not been designed to fit the needs of WBL and are based on outdated technology. Open source tools and libraries which supply state-of-the-art machine learning could solve the existing needs. The challenge would be to find the right constellation of data, algorithms and business goals. In addition, it has been mentioned that the use of AI-powered products could work efficiently in an agile approach such as SCRUM which can be implemented within the CRISP-DM framework.

To minimize the risk, it is suggested that the PCP could be split into phases of: Proof-of-Concept, Minimal Viable Product and Productive Code. The TCO could be lowered by intertwining development and delivery.

It has been suggested that using a cloud provider for storage of collected data from different systems could simplify the data ingestion and processing and potentially lower the cost of delivery. The adoption of a dedicated IoT platform in a multi hybrid Cloud environment would enable WBL to reduce Time & Capital spent on this and subsequent technology developments. Integration with proven and novel technologies deployed in a hybrid cloud environment is also mentioned as an option.

Another suggestion is to include only those parts of the network around Venray in the PCP which are equipped with IoT solutions that provide enough historical data, leaving the unmonitorable parts of the network outside of the PCP.

A list of failures or system states that WBL wants to observe could be used to focus on creating a model that predict these kind of failures.

The use of Big data step by step could avoid spending big sums of money before obtaining actual feedback from the application. Also the reduction of data storage, reduction of data communication and generation of models based on historical and actual data instead of more expensive mathematical models, could lower the costs.

"Start Small, grow big" is also proposed as a means to select a scalable solution that can grow as more sensors are connected would provide more control over the budget. Thus, separate the analytics software and the sensor hardware into separate tenders. The use of a boxed approach could keep the solution manageable, flexible and interchangeable. One use-case as a closed and self-containing box that uses its own components can assure that the whole system keeps functioning when boxed-components are replaced for different use-cases. This can make it possible to remove use-cases from the landscape lowering costs and avoids vendor-lock-in.

A Brainstorm/risk session could address the following questions:

- Is the business case / value clear and set up a good definition of value.
- How to robustly incorporate ML models into your work process?
- Is the deployment of ML models possible at the edge (on site) or also cloud based?
- To what level of complexity will the ML solution interact with the user? (visualisations / alarms / advice / control)
- Is there a clear description of existing and non-existing data relevant to the scope?
- Can additional low cost sensors be used on site, and how are these interfaced in (near) real-time to the models?
- What is the approach to work together (ideally data science team close to domain experts, leading to iterative creation of MVPs)
- Is centralization of data prep and cleaning possible between all parties; this often is 90% of the workload when creating solutions.
- Are the models created as bespoke models, or are they based on pre-existing models?
- Are the models created and deployed using a platform, or are they deployed in a custom manner?
- Avoid potential costly and lengthy process of agreeing standards between all parties by centrally determining one acceptable standard, including all datasets. Support all data delivering parties by for example developing ETL tooling for data conversion to the standard. Responsibility of complying to standard lies with data delivering parties (e.g. municipalities and their

contractors)

The scope of the project and requirements would determine the most appropriate supplier and tools to be selected. The focus of implementation should ensure that all the tools envisioned are actually required. A definition of where the IoT sensors are going to be placed and the analysis of the existing network is important for the deployment which is expensive to perform and maintain.

Definition on the requirement of a real-time need to monitor performance or a short term selection of priorities or a medium term planning need would be needed as each of these may sound similar and overlap but will require different tools and activities. In this regard, the contextual asset knowledge or understanding of using analytics specifically for the water industry is of importance. A possibility could be to have a supplier that can provide proven predictive models, in a pre-packaged solution, and only perform Research & Development where it is needed as this is more expensive than a simple implementation of existing predictive methodologies.

Issues such as sensor maintenance, disparate data protocols, sensor data quality issues, closed architecture software packages, lack of understanding of the sewage system, lack of knowledge of operational limitations, and lack of understanding of the desired business outcomes can result in redundant work, rework costs, misaligned expectations, and reduced business value.

A successful and cost-effective delivery of a Big Data solution would need to consider a partnership cooperation; experience in implementing IoT, Big data and digital technologies for sewer systems; and an open architecture that leverages past investments in software, information, infrastructure, sensors, etc.

Another suggestion is to use o a Real Time Decision Support System (RT-DSS) framework and approach in order to 1) contextualize, network and learn from the data; 2) use the data to train the utility's models with AI applications; 3) transform the data and models into real time operating systems that allow WBL to understand past events and current circumstances, as well as predict the range of near-future outcomes, so WBL can automate work orders for predictive maintenance and asset set points.

Other suggest using an off-the-shelf product for big data management and analysis such as WeRiot that offers a full-featured data management system which combines aggregation, import/export, storage, analysis and visualization of data into one secure, flexible platform.

In addition, to prevent failures of the PCP it is suggested to include minimal isolated prototyping of possible sub solutions very early on in the process during Phase 1.

The provision by WBL of asset system characteristics (e.g. pipe diameters, pump capacities etc.) in standardized formats (through data standards like for example DAMO and GWSW) would enable to automate the setup/configuration of the Big Data solution saving time.

Software as a Service that incorporates off the shelf trained algorithms for specific use-cases, which allows to incorporate new algorithms for new use-cases is also mentioned as a possibility.

Finally, other suggestions are the mandatory use of free open source technology; better (detailed) specification of requirements / targeted business value, categorised and prioritised (possibly in MoSCoW methodology); the use of standardisation; the use of scalable architecture for the big data platform, reducing initial costs. Platforms such as AWS/Azure/Google etc. would integrate with their other services (AI / ML modules) and are easily scalable; and no vendor lock/custom software.

2.2. Appropriate budget for delivering the service

The budget estimations range from 150k-220k, 400k-600k, up to 2~3 million Euros total, for Phase 1, 2 and 3. Some parties suggest different budgets per phase (e.g. phase 1: 15K, 40k or 50k-, phase 2: 35K, 85K, 120k; and phase 3: 50k, 100k, 150k).

2.3. Suitable technology for Big Data in the field of water management

Some parties who replied to the questionnaire mentioned to have experience in the field of water management but other have experience in areas like predictive modelling but not in the water sector. For example, respondents have indicated extensive experience in using machine learning to obtain actionable insights from IOT-based sensor data from concept to productive environment; knowledge in the area of predictive modelling, time series, analysis and optimization focussed on real-time (streaming) analytics and monitoring with huge volumes of data to provide actionable insights and alerts.

Respondents also indicate to have developed a technology that draws on mathematical models built with limited data on rainfall and water levels to create flooding predictions for rivers.

Parties refer to knowledge around Big Data Solutions/Analysis for Water and use cases. Some have experience in the water industry, other have developed a mathematical model that can find optimal parameters when machine learning is used to train the model with past rainfall and water level data, creating functions based on the tank model concept, which expresses water discharge from a river basin within hydrology.

It has also been mentioned a stack of Big Data analytics solutions utilizing real time decision support systems (RT-DSS) that implements the sense-predict-act framework. It has also mentioned proven knowledge and experience in Big Data Solutions. For specific water management domain expertise an ECO-system approach where multiple parties need to work together in either a partnership or consortium model has been proposed.

Another technology mentioned is the Smart Industry Assistant (SIA) which is powered by SAS technology which can process real-time data and provides standard modules.

List of technical solutions:
a. predictive maintenance
b. prediction water quality / water flow / etc
c. alarms (e.g. pump failure, to reduce unplanned maintenance)
d. level prediction based on weather and system performance
e. performance estimation
f. control of additions to minimize turbidity (automated / advice)
g. cost optimisation (maintenance planning)
h. allocations of resources (adaptive / request based control of system)
i. early leakage detection

Method of deployment:
a. automated model deployment and API creation tools to minimise the cost of taking models into production
b. cloud based development environment / computation environment to enable elastic scaling and hence reduction of costs at time of deployment/production
c. interfacing and converting data formats of real time external systems using pre-defined procedures and drag/drop workflows.

2.4. Suitable methodology/approach to estimate the impact of solutions

Diverse approaches to estimate the impact of solutions have been mentioned, for example: a scalable platform for wastewater operations; qualitative and quantitative analysis; to look at the historic network outages to compare past incidents and overall costs associated with historical failures; benchmarking performances of WBL and other water boards; measuring current performance and cost against a suitably defined KPI informed by policy; end-to-end view and intuitive ways of visualizing and understanding patterns and anomalies; Deep Design Thinking; benefit analysis of doing nothing or taking a traditional approach; data modelling techniques that allow for predictions and better control of the water and sewage systems; automated performance assessment; benchmarking through the analysis and identification of the characteristics of a specific use case and estimation based on comparable (historical) use cases the impact; start with a Proof of Concept and discover the actual impact; and the creation of business cases estimating the impact of solutions and cost savings.

2.5. Classification for potential technologies

Some parties have suggested as classification for potential technologies Agile software development and Lean by design. Other suggested that the classification can either be seen as a have multiple states, such as: fail, all-is-well or partial failures which an algorithm could distinguish. It was also mentioned a value stream design made with the stakeholders, WBL, the public (users) and the technology Provider based on pay per use, pay per functionality, pay per performance.

A number of areas was also distinguished:
- EAM/ERP
- Financial Systems
- Dashboards & Control Panels
- IOT Sensors/SCADA
- IOT software
- Big Data
- AI/ML Predictive
- AI/ML Prescriptive
- Projects Portfolio

Other suggestions point at a Human Centric Co-Creation approach that focusses on the added value and business benefits for the organization that implements it. Digital Co-creation is referred to as a way to leverage digital technology and deliver innovative value jointly with partners and customers.

Other potential technologies could be broadly classified based on the RT-DSS framework: sensing technologies (including those needed to transport and store data), prediction technologies (including machine learning, digital twins, and data assimilation), and act technologies (including visualizations, workforce management, and pump station management). Top-down approaches are mentioned as generally generating architectures that are cost effective and successful.

It was also mentioned the trend of combining different technologies to create new applications, such as automated driving will not be possible without 5G and IoT. The data-driven society hinges on technologies such as the Internet of Things, blockchain and artificial intelligence.

Other mention MoSCoW and SWOT strategies, AI knowledge Map, and heuristics, Physics-based models and AI-based methods.

2.6. Approach to developing and testing technologies

The approaches mentioned include: Agile development methods to minimize risks, cloud platform, CRISP-DM, AI/ML testing, Model Factory framework, WeRiot's Catasta platform, Design Thinking, GAP analysis, a stepped approach where the solution is first tested in a model (either a conventional Hydraulic/Hydrologic model, a mathematical model, or a neural network) which demonstrates the probability of achieving the KPI. Testing is also suggested to happen at the Venray prototyping and a scrum way of working.

2.7. Barriers to use existent solutions in different settings

Some possible barriers mentioned are:

- Building data pipelines to fit existing solutions could lead to overhead.
- Non-mainstream machine learning and AI libraries may fall under licenses such as GPL where one cannot use such libraries without making that portion of their code publicly available.
- If the type of pipes or the structure significantly differs from the network in Venray a reevaluation must be made on how applicable the created models are.
- IP related solutions with increased learning curves
- Selection of location for IOT sensors is critical as they are expensive to deploy & maintain. They generate a lot of data that may be very useful or may be noise. Overuse of sensors will create a huge data burden that may not provide any benefit to Limburg.
- Assessment of the quality of data, understanding of customer rules, pre-

existing information governance are all aspects to be considered prior to building machine learning models and/or physical models. Also, deployment is dependent on the result of Lot 3 Big Data which is an unknown at the moment.
- The type of data that is collected determines what is possible to detect.
- Different settings could result in different behaviour of the systems which could extent the learning time before insights/predictions can be generated.
- Standard products are great for standard problems.

One sees only opportunities.

2.8. Barriers to implement Big data solutions

Some indications are:
- Zero-trust approach to software design
- If data cannot leave their on-premise storage, or that a key part of it may not be used due to privacy issues, this may slow the implementation process and make the task of ensuring the security of the overall system more enduring.
- Assure that (good) historical data is present, known / logged faulty events, malfunctioning equipment.
- For maintenance it is important that the solution is maxim standardized, has a long term (10 years+) lifecycle support, has no performance & storage limits in major upscaling of the objects in the field. For Cybersecurity, the solution should be 'secure by design' and a joined approach together with the company which is executing the Lot 2 (cybersecurity) part of this tender.
- Security will need to follow government requirements for Critical Infrastructure ensuring a secure layer to prevent the IOT sensors being an attack vector.
- Placement of IOT sensors below ground presents issues for connection to the network and maintenance, which increases the cost.
- A barrier is determining if the right problem will be solved in the proposed areas of maintenance and security that also delivers a positive business case.
- Suitable access to SCADA monitoring and control system data, operational control system and infrastructure design and performance data.
- The combined use of external and internal data can be an issue. The current Lot 1: Process Automation and Lot 2: Security are seen as unknown factors at the moment.
- A key aspect that could create a barrier to implementation are mismatches in specifications between the different layers required to implement the

solution(s). These specification mismatches can occur when disjointed groups implement different elements of the Big Data solution. A typical example is a mismatch between the delayed or spaced-out sensor transmission times and the time horizon for making real time decisions.
- For the very large scale machine learning projects it has been found that in order to be able to rely on them, customers need to absolutely understand the detailed functioning of the models that are used to drive the solution.
- Standardisation issues, it is possible that not all equipment is suited for sharing data (at this moment).
-

2.9. Relevant functionalities not described in the Market Consultation Document

The following aspects have been pointed out:
- Apart from direct monitoring, look for information hidden in the meta-data of the pipe network. Regression trees on this information can support direct maintenance efforts to those spots which are most likely to fail.
- The use of a transparent way to account for the origin and authenticity of data coming from different parties as well as the processing of this data in the automated pipeline, through a private blockchain structure.
- ML model production platform which is independent of (and integrable with) any sensor and control software vendors
- Pre-defined models for water-focussed AI scenarios based on research to date and which can be optimised/tailored easily.
- Flexible integration with IoT telemetry and control systems
- Full data (format) transformation and real-time interfacing tools using a highly performant and drag/drop UI and maintenance capability based on common data standards
- Real-time monitoring of the system, including data collection via image recognition.
- Daily/ weekly analysing the performance of the system.
- Short and medium term operational and maintenance interventions on the system.
- Strategic intervention on the system (Replacement, renewals, different usages).
- Hyperpersonalisation, Image recognition, Autonomy, Anomaly Detection, Prediction, Conversation and Goal seeking. The ML functionalities of the proposed system can best be described with reference to some of these general use cases e.g.
- Including a greater variety of external data sources would help the effectiveness of the system. Weather is one but perhaps larger scale social

conditions could have influence through changing demand (holidays, time of year, and variation during a week …). Direct connection to staff availability can be integrated to the planning of predictive maintenance and preventative inspections.

- Graphical data science and predictive analytics platform designed for users of all skill levels to deploy insights at scale to improve their business.
- Full-featured data management system which combines aggregation, import/export, storage, analysis and visualization of data into one secure, flexible platform.
- Digital twin to verify the Big data solution with each software or configuration update that could be used in Phase 2 of the PCP to validate the prediction. It can also be used to simulate extreme circumstances and the reaction of the Big data solution to these circumstances.
- Integrate weather forecast in wastewater transport models as well as implement real time control (RTC) of pumps to optimize the performance of the transportation network as a whole. This can be used to eg. manage combined sewer overflows (CSO).
- Self-service dashboarding and self-service analytics
- Scenario analysis. Once a good model is established, a wide range of questions can be answered. For example: what happens when valve 'x' fails? Or, if there is a leakage in section 'x', what would the current set of sensors record at an early stage? Many more questions could be considered.
- Flexible system modelled properly into sub-modules, with good understanding of the dependencies between these modules. A fully black-box machine learning approach may not be suitable for such an application. A solution based on physics and engineer knowledge, augmented with data-driven solutions (machine learning) would be more suitable.
- Advanced dash boarding (applied at other central control rooms, facilities). Good knowledge which data is important to display and act on.

2.10. Inadvertencies in the requirements

The following comments have been provided:

- The requirements were not written with software development in mind. This should be addressed urgently.
- A current technical state of the system is missing. It is unclear what kind of sensors are deployed, and at which part of the network. Therefore, it is challenging to estimate which type of analysis can be conducted. We propose to gain additional insights in the type of sensors, there placement and what kind of data (variables) they report.
- The scope can be clarified further.
- The requirements may be fulfilled with a smaller and/or simpler set of

components than outlined within the Market Consultation Document. Reviewing the requirements with the Limburg team may yield further insight to reduce the delivery risk of the research project and by focussing on application and integration of existing technologies before creating new solutions.

- It would be helpful to see the actual data that is coming from the source systems and how this data is transported to its destination. Also the time it takes from measurement to receiving the data is vital information to determine the possible use cases for a solution like this.
- More insight into the available data and the patents concerned could be helpful.
- Opportunity to leverage the technologies implemented with other parts of the urban water cycle such as treatment and distribution systems.
- Opportunity to elaborate on the challenges that the utility is addressing so as to better configure or customize the solution to cost effectively address the problem.
- The question is specified in terms of the methodology that needs to be used ("Artificial Intelligence models …"). However, the real question is how to effectively (automatically) manage the complex water system (operations, maintenance etc.). Would WBL mind if no AI was used at all, but the water management runs smoothly? In other words, AI should not be a goal in itself. The PCP should start by first understanding which practical problems must be solved by the solution. Allow the market the freedom to propose the best solutions. It is likely that the solution includes AI to some extent, however, positioning the question with too little room for (partial) alternatives does not seem most beneficial.
- Heuristics and physics-based approaches aren't mentioned in the published market consultation document.
- With the current proposal of phases, it wouldn't be possible to know if the IT architecture would be suitable to solve the real problem at hand.
- The proposed timeframe is rather short to really go from new ideas to implementation. To provide meaningful innovation rather than finetuning software that has already been made, more time would be needed.

"

2.11. Recommendations regarding the PCP

Among the recommendations are the following:

- Modular design is appropriate to try and test different algorithms and approaches.
- The automation of cognitive task load management-- assuming that those who are in charge of monitoring dashboards and are responsible for making decisions need to be at a satisfactory level of alertness at all times, the

automation level of certain tasks may increase or decrease to ensure the human with decision making would remain in a suitable alert zone.

- Widely inform employees in what a predictive tool can predict, and moreover explain its limits. This minimizes the risk of a “computer knows best” mentality. Furthermore, it makes it clear which states of the network are automatically monitorable, and which parts require human observation.
- In the future, the solution should be able to be extended with complete different kind of data, like: audio, pictures, text and video. Including suite analysing and AI tools for these different kind of media.
- Include data quality assessment in the process. Adding this step to the process has proven to be valuable. Furthermore, algorithms could be included to determine the condition of the sewage system based on image recognition.
- Connect other technologies besides Big data to this solution like: Blockchain – for sharing IoT data with 3rd parties that deliver services to the whole chain. Quantum Annealing –to quickly optimize the water flow calculations to find the optimum. Robotic Process Automation – for automating the software that surrounds the solution. IoT Asset Management – for geo-fencing and maintenance planning. Visual AI – using drone / satellite / video footage to monitor the status of rivers, water levels, and equipment in an automated way. Critical Infrastructure Security – Using geo-fencing and asset tracking to secure the critical areas and monitor in real-time what comes in and out of (e.g.) a water plant. Automated Sewage Inspection - Using automated camera inspections to detect issues within the sewage system automatically.
- Apply integrated People, Proces en Technology approach.
- Take an explorative approach with regards to data and make the invisible visible. In many cases data exploration leads the way to unexpected value.
- Focus on all data:Real time, streaming and historical data. All data types: structured, semi-structured and unstructured, Big data and fast data.
- Focus on reliable, fast and secure information exchange within the water chain ECO-system.
- Include exploration of Digital Avatar concept in the PCP.
- “Think big, start small, learn quick and scale fast”.
- Join the National AI coalition in The Netherlands.
- Avoid vendor lock-in and make it possible to use “best of breed” tools that are best suited for the job. There is “no one size fits all” .
- Collect all relevant historical data, possibly from other similar entities, using hardware with open protocols might be a good idea to avoid vendor lock-in and ensure interoperability.
- Big data solutions require good understanding of its functioning within the

company of the end user. Therefore an internal (project) group at WBL should be established to function as stakeholders during the PCP and who will take daily control over the Big data solution in cooperation with the contractor for Lot 3.
- Reconsider the unlimited royalty free non-exclusive license for indefinite time. There are doubts of whether an unlimited royalty free non-exclusive license for indefinite time is fair. Especially if a significant part of the requested functionality is already functional on a specific platform.
- There is a risk in trying directly the most advanced insights/predictions, resulting in disappointing results. Investigate what the most valuable insights and predictions are to ensure focus on the aspects where it is expected to get the most value. Start small, but think big. Use a modular setup of the solution to get value of the initial modules and build from there to more advanced solutions.
- It is important to get WBL end-user acceptance of the to be developed functionality, to ensure its adoption, so it should be inquired about envisioned strategies on how to achieve this goal.

2.12. Relevance of patents and standards

The replies indicate that:
- Insufficient depth is provided, but the outline of the analysis seems correct.
- Although patents appear to be relevant, given the relevant recency of topics such as big data, these patents may not cover all required aspects when it comes to a domain such as water management.
- several best practice approaches and methodologies that have been deployed within the water industry are not patented. Most ML/AI techniques are public domain, however their application to the water industry requires contextual knowledge not available to all vendors.
- Some have patents pending for the developed AI Data Analytics algorithms.
- Some more detailed information is required about the patents.
- Every patent and standard must be considered whilst designing a solution. Not only during phase 1, but also in sequent phases 2 and 3 there should be a check by IP-experts on whether or not the designed solution infringes third party intellectual property rights (or licenses for use of IP-rights have to be purchased). Relevancy of the listed patents and standards in the prior art analysis must be determined in the course of the project. It is convenient however that a general analysis on patents and standards for this topic has been already performed and shared.

2.13. Additional relevant standards and patents, other rights and trade secrets

Some participants mention no other relevant standards and patents, but indicate that AI algorithms and prediction models would fall under patents or trade secrets. Usually an usage license right can be purchased for the usage of these components. Other participants refer to a number of Publications and Standards/Patents which cover the scope of the project. For example, the Jefferson Project which concerned the Smartification of Lake George in New York (USA). IoT-based sensors are being used at Lake George in order to collect data that can be used to improve the freshwater ecosystem. in order to monitor threats such as pollution from road salt, invasive species and excess nutrients. The Jefferson Project began in 2013, and it includes a smart sensor network that gathers more than 9 terabytes of data annually, and it also uses computer modeling to better understand freshwater ecosystems, and provide science-based insights.

The RAMI 4.0 Reference Architectural Model and the Industry 4.0 components are mentioned as giving companies a framework for developing future products and business models. RAMI 4.0 is a three-dimensional map showing how to approach the deployment of Industry 4.0 in a structured manner.

Finally, it is mentioned that the platform will most likely be built on software that is subject to IP-rights of third parties (e.g. Microsoft). It is foreseen, depending on the nature and concrete development of the solution, that certain licenses have to be purchased. In the course of the project it should be possible to identify these in more detail.

2.14. Ownership of Intellectual Property Rights

While several respondents indicate no ownership, some indicate that Microsoft and other own relevant IPR, platforms and models, other seem to own components, AI models and software that are part of a potential solution designed to automatically switch to a newly provided data processing program when a parallelized data processing job has been completed, by separating stream processing into data reception processing and actual data processing so that data reception processing and current data processing are not stopped.

Other participants indicate to own IPR with regard to a software suite that provides a significant part of the expected functionalities that WBL is looking for. Both for transportation systems as well as solutions for the waterchain in general (transportation systems, wwtp smart control and sludge transport and dewatering).

2.15. Patents that may constitute a barrier

Mostly no barriers in delivering the solution envisaged in the PCP have been identified. It is also mentioned that during phase 1 of the project it is difficult to determine whether or not any patent constitute a barrier to delivering a solution in the envisaged PCP procurement. Potentially every patent can become a barrier if the third-party owner of IPR is not willing to provide licenses (against a royalty fee). In principle, third party IPR will mainly be a cost increasing factor due to the royalty fee, which could potentially also be a barrier.

2.16. Suggestions of a more suitable list of developments

While several respondents have no suggestions, some think there are relevant and interesting developments that would be very useful for WBL like: Blockchain, Quantum Annealing, Robotic Process Automation, IoT Asset Management and Visual AI. For all these technologies, it is suggested the creation of a suitable business-case should always be executed to ensure a suitable usage of the new technology.

Certain developments in this particular field concern sensitive information which cannot be shared with other parties involved. In addition, since almost every day new patents are submitted and acknowledged, it is needed a close attention of the development team, prior to delivering a solution, to ensure no third party IPR is infringed.

2.17. Commercialized Big data solutions

Most respondents indicate no commercialized solutions in the water sector but refer to other industrial segments. Other information is presented as confidential.

Reference is made to comprehensive framework approaches to water chain automation, maintenance and security, large system wide sensor networks throughout the water chain – from water intake to tap and collections network to effluent pipe – to collect and analyze data using both traditional and machine learning methods. It is also mentioned decision support systems that provide feedback, situational awareness and operational recommendations via user-in-the-loop dashboards as well as automated real-time controls.

The "platform-as-a-service" offering based on “pay-per-use” models are referred to, as well as a general-purpose data management platforms to use in either industry and either scope of application. The platform includes solutions for automation, maintenance and security.

A Smart Industry Assistant powered by SAS technology is said to process real-time data and standard modules that allow to:
o Self-service analytics/dashboards
o Monitor effectiveness of systems
o Detect anomalies in normal behaviour of systems (e.g. to detect leakages, clogged systems)
o Predict the performance of a system
o Smart Maintenance (including predictive maintenance)
o Connect with maintenance systems
New modules/algorithms can be added and customized. Furthermore SIA is for example used to monitor pumps in real time and detect anomalies and predict future events.

3. Prior art analysis and outcome of the preliminary market consultation of 2018

In 2018, an analysis of prior art in the form of protected intellectual property (patents and literature) and technical consensus (standards) was performed. The prior art analysis was performed with the aim to gain insights into the available technology and methodologies regarding Big data solutions for the water chain automation, maintenance and security, to identify the active and innovative market players and to inform the tender documentation in the follow-up procurement.

Based on keywords derived from internal project documents and project partners, it revealed 170 potentially relevant patents with unique technical content, 19 relevant scientific articles and 72 relevant standards (see table 2 below).

Regarding Big Data solutions, there appears to be room for European suppliers to innovate, and incentives for them to do so, given that there are almost no relevant patents with coverage in a Dutch or European wide innovation procurement (1.2% of total)[32] and no patents with global coverage. There are no patents belonging to a patent family (i.e., receiving coverage at more than one patent office), and the patents are relatively dispersed among a number of suppliers (the company with the highest number of relevant patents has 5 of the 170, or only 3%). The vast majority of patents are held in China, and by Chinese companies. Compared with patents in Phase 2: Security, patents for Big Data had much higher Technical Relevance (TR) scores (18 patents with a TR > 3), indicating that there are a number of patents for which certain attention must be given as they are more technically relevant to other solutions in the area. These patents are held in China, Japan, and the US, and therefore while they do not block similar innovation in the Netherlands or the EU, they may have technical characteristics which can benefit suppliers developing new solutions while having a low risk of being infringed.

Literature for Big Data is primarily on methods and models for water treatment systems, including foci on chemical analysis, prediction, and monitoring. This is often based on statistical analysis, and offers early-stage research that may be of interest to WBL in defining the solution space or supplier requirements. Standards for Big Data have been set by industry consortia (e.g., NEMA National Electrical Manufacturers Association) and standards-setting organizations (ASTM, NEN and DIN, among others). Topics range from increasing energy efficiency, to automatic sampling, to field bus systems and control / automation. The 2003 European standard was harmonized by countries including by NEN, DIN (Germany), PN

(Poland), and SN (Slovakia), entitled: "Wastewater treatment plants - Part 12: Control and automation". International standards cover information systems connections in local area networks (ISO).

Table 2: summary of findings of the prior art analysis

Document Type	Item	Results
Patents	Total	170
	Jurisdictions Patents held in jurisdictions relevant to a European-wide project: EU/WO (% of total)	2 (1.2% of total)
	Top 10 Patent Applicants (by TR)	• Beijing Tsinghua With Scale Planning Design And Research Institute Co. Ltd. • Bridgestone Corp. • Hu Xiao-en • Li Jian-guo • University Sichuan • Xu Qiu-tao • Xing Xiao-jun • Guangzhou Chime-long Water Supply Equipment Co. Ltd. • Zhengzhou Water Industry Science And Technology Development Co. Ltd. • Ceng Chen-lin
	Suggested codes for choosing Common Procurement Vocabulary (CPV)	73100000-3 Research and experimental development services 73110000-6 Research services

Document Type	Item	Results
Literature	Number of articles found	19
Standards	Number of relevant standards	73
	Standard-Essential Patents (SEPs)	No

	Represented Standards-Setting Organizations (SSOs) with related standards	• GOST R (Federal Agency on Technical Regulating and Metrology, Russia) • Finnish Standards Association • DWA (Deutsche Vereinigung für Wasserwirtschaft, Abwasser und Abfall e. V.) • Verein Deutscher Ingenieure • ASI (Austrian Standards Institute/Österreichisches Normungsinstitut) • Verband Deutscher Maschinen- und Anlagenbau e. V. • DIN • CEN • DVGW (Deutscher Verein des Gas- und Wasserfaches e. V. - Technischwissenschaftlicher Verein) • ASTM (American Society for Testing and Materials) • FGSV (Forschungsgesellschaft

Subsequent to the Prior Art analysis, a market consultation was organized in 2018. The market consultation of 2018 had the purpose to collect input from market parties on the procurement of the 3 Lots (Process automation, Security and Big Data Storage). In the context of this market consultation, preliminary input has been received on the technological possibilities and approach to the Big Data analysis, which forms the subject-matter of the PCP.

Hereafter, we present an anonymized summary of the most relevant input received from the market regarding the Big Data storage and analysis.

As Big Data is a rather new area in the IT field, the pace of technology development is quite high. Thus, still few standards have arisen in this area, and that is confirmed by most of the companies in the market consultation. That is particularly true for the Big Data analysis, that would seem the most challenging to standardize, since is mostly driven by the final users' specific requirements and needs (e.g. predictive maintenance).

A point that should be emphasized when looking at potential standards and solutions for Big Data is to make a clear distinction between *data storage* and *big data storage*. The former is not new in the IT industry, and there are indeed plenty of solutions available in the market for corporate needs. It follows that there are many standards related to it. The second is the new frontier of data and just few standards seem to be available.

Some market parties argued to have created their own platform to collect and analyze sensor data, although a patent has not been pursued.

Based on recurrent comments, it would seem that any kind of Big Data initiative (especially from a data analysis perspective), before even considering potential corporate solutions, would require a structured and efficient planning. This preliminary phase can be tackled through various standard methodologies, which may be considered particularly in the PCP.

Several market parties proposed to have a reference **Hybrid Data Management** for grounding and validating the decisions based on the Big Data. This system is a standardized approach on how to manage and take decisions in complex data environment. Moreover, it would serve as a preliminary exercise to choose the most suitable Big Data solutions.

Market parties also requested clarification from WBL on the following aspects:

- Functional requirements, including purpose and intent per functional component
- The components to support software and service selection processes that are required for setting up a platform.
- Strategic design principles that serve as guidelines for the market party in defining detailed design and implementation
- Determine usage patterns - the merging of similar use cases and applications - and relate them to the organizational value and priority. This way a clear roadmap can be determined for execution of projects on the platform.
- Define key conditions for data flow patterns early, for example which data sources need to be included - so that timely measures are taken to determine the availability and the input of data
- Non-functional aspects of components – e.g. whether and which cloud strategy should be used applied - to ensure optimal deployment in terms of cost efficiency and business value
- To gain insight into which expertise and knowledge capacities are required and how these are being invested for the realization and operation of the platform.

The reference architecture therefore does not have technology or standards as a starting point, but is the guiding principle and demands future stability and modularity.

In other words, when a technology no longer satisfies the wishes of WBL, it can be phased out, without the reference architecture and the platform suffering from it. This gives stability and control, whereby initiatives from PCP can also be fitted.

A similar approach for the design and implementation of solutions based on big data analytics could be based on the **Cross-Industry Process for Data Mining (CRISP-DM)**. The CRISP-DM is not a standard per se, rather a structure to plan a data mining project, based on a number of data mining phases. In market parties' view the CRISP model is the most used standard approach for Big Data planning.

The following steps are hereby relevant:

1) *Business Understanding - building a good understanding of the business needs, definition of the research area and concrete research questions.*
2) *Data Understanding - building a good understanding of available data, including analysis of relevance and usability for the research question.*
3) *Data Modelling - structuring, cleaning and enriching relevant data for optimal application in an analytical model.*
4) *Analytics Modelling - applying analytical models to the prepared data in such a way that insights are obtained that help to answer the research question.*
5) *Analytics Deployment - answering the research question through visualization and presentation of the obtained insights, including in some cases operationalizing the analytical model in a continuous improvement process.*

The market parties have suggested open source solutions for the big data storage (such as *Apache Hadoop*).

Market parties have also shared their views on technological developments in this field:

- *The making of analytical models often takes place in the languages R and Python, to a lesser extent we encounter Matlab.*
- *For the good management and reuse of analytical models in larger organizations, development environments such as Rapid-Miner and R-Studio are widely used.*
- *Most commercial suppliers of reporting and visualization products now offer good opportunities for integration and visualization of analytical models in a managed and secure environment.*

As said, Big Data is a fast-developing sector and at least from the point of Big Data Analysis the current market trend is varied and difficult to define. Still, there are new trends that suppliers are confident to see implemented in the near future (if not already) such as **Artificial Intelligence** and **Machine learning**.

Finally, one market party pointed out that "*real-time processing, decision models and machine learning require rapid switching that the infrastructure and data sources of WBL need to support these functionalities. In addition, the question is which intermediate tools you will use to convert data and make the calculations*".