

## A Simple and Reliable Unstructured Data Synchronization Technique for Biogas Data

KKLB Adikaram<sup>1,2,3</sup>, MA Hussein<sup>1</sup>, M Effenberger<sup>2</sup> and T Becker<sup>4</sup>

<sup>1</sup>ResearchGroup of Bio-Process Analysis Technology, Technische Universität München, Weihenstephaner Steig 20, 85354 Freising, Germany

<sup>2</sup> Institute for Agricultural Engineering and Animal Husbandry, Bavarian State Research Centre for Agriculture, Vöttinger Straße 36, 85354 Freising, Germany

<sup>3</sup> Computer Unit, Faculty of Agriculture, University of Ruhuna, Mapalana, Kamburupitiya, Sri Lanka

<sup>4</sup> Lehrstuhl für Brau- und Getränketechnologie, Technische Universität München, Weihenstephaner Steig 20, 85354 Freising, Germany

### Abstract

Data from biogas plants are considered as unstructured data due to the complex and dynamic nature. Thus, it is not possible to define common data structure for data generating from biogas plants. This reason prevents online or inline synchronization of biogas data with the central database. In addition, due to the scattered locations of the plants, it is not possible to use private communication media for data synchronization. Also, use of existing complex fully automated systems with high technology is not economically feasible. We introduce a software solution that is capable of overcoming aforementioned drawbacks. Total system is a combination of data recording module and two data synchronizing modules. The data recording module of the system is installed in a personal digital assistant (PDA) which is capable of storing and locally customizing. After daily data recording, connecting the PDA to the first synchronization module in computer will automatically update new data to a spread sheet in the local machine and sends a copy of new data to central computer as an email. In the central computer the data in the email are uploaded to a database using the second synchronization module. This process guaranteed continuous data flow of unstructured data to central database at low cost while providing farmer much reliable data without additional effort. The proposed method has been implemented for synchronizing biogas data of six plants with different data structures that located in considerably large geographically area (Bryan, Germany).

**Keywords:** Biogas, Data Synchronization, Unstructured data

### Introduction

Most of the data generated from biogas plants are recorded automatically using data loggers. However, some of the data like daily material feeding cannot be recorded automatically. The normal practice is to record them manually (using a data log book) and subsequently enter such data to a local database. This is a time consuming error prone task. Entering previously written data to a database can be one of the most possible error prone situations. This is a duplication of work and consumes extra time and money. When considering the physical structures, biogas plants have many different types of designs and produce huge amount of heterogeneous data. Hence, data collection from biogas plants is considered as a complex procedure. In addition to the data produced in the biogas plant, there are considerable amount of data from experiments as well as from analytical laboratories. This makes structure of biogas data that doesn't have a pre-defined data model, which is known as unstructured data (Raj and Deka, 2014). Thus, it is not possible to create a static structure for biogas data, which can store all the data related to different types of

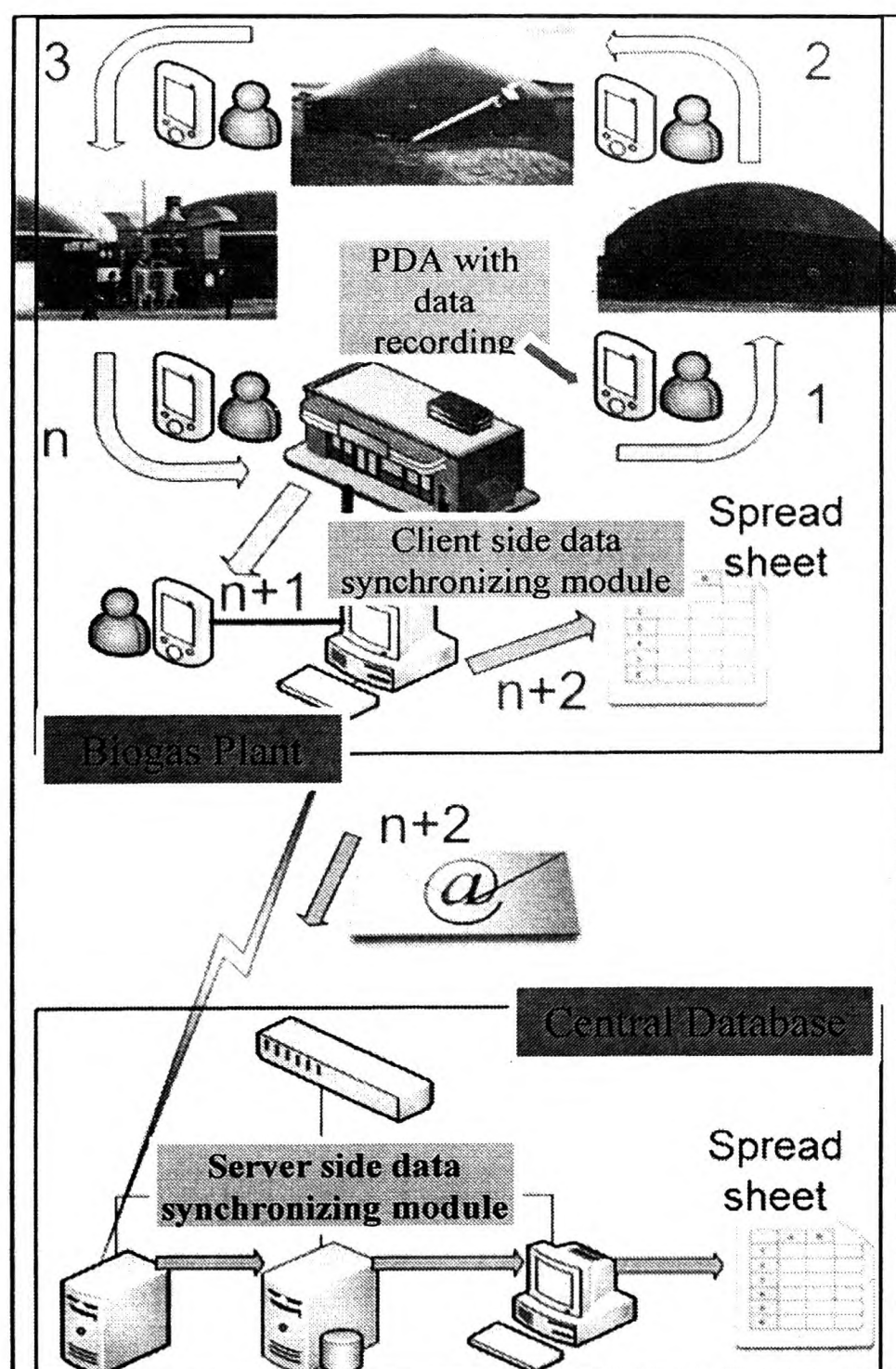
biogas plants and experiments. In the process of monitoring and evaluating performances of biogas plants with different structures and technologies spread over large geographical area it is a must to synchronize all the data of each plant with a central database. However, unstructured nature of biogas data does not permit to create a common data structure that is suitable for all the biogas plants (Inmon and Krishnan, 2011). Because of that, it is not possible to synchronize data of remotely located biogas plants with a common database (Inmon and Krishnan, 2011; Sukumaran and Sureka, 2006). In the process of data collection, communication cost is an important factor (Golubchik *et al.*, 2013). When the data collection points are located and spread over a large area of a city, town or a country it is difficult and financially not feasible to establish private communication channels between collection points and the central location. Then the solution depends on the public communication systems.

The most common public communication mediums are telephone, email, internet services,

and cellular networks. Using email for sending data is a reliable and low cost technique that can use any public communication media.

### Materials and Methods

In a biogas plant, locations of data capturing points are located in different places, which is again depend on the physical and technical design of the plant. To replace the data log book, we decided that the most suitable device is a PDA. A laptop, or tablet PC is not suitable especially because of the physical size of those devices and cost. Even though laptop is a portable device, it is not a convenient device for entering data in different places, sometimes while working. The developed method consists of three components, namely data recording module, client side data synchronizing module and server side synchronizing module (Figure 1).



**Figure 1:** Data capturing and data synchronizing procedure of the proposed method for a single biogas plant. Number shows the data capturing/synchronizing sequence at the biogas plant.

Data recording module was designed for PDA, which is capable of storing required data. Since the biogas data are unstructured, the module in PDA was capable of adding, removing, and modifying variables related to the biogas plant. Client side data synchronizing module is designed for the PC installed in the premises of biogas plant. The main role of this module is to send the collected data to the central database as an email, while uploading the same data into a local data base according to the required format by the farmer. We used spread sheet as the client side database because of the very simple nature of it. Server side synchronizing module is for reading the emails from different biogas plants as emails and extract the data in the emails. After extracting data in emails, these data were appended into different spread sheets which belong to individual biogas plant. Structure of each spread sheet is unique and follows the data structure of data from a biogas plant. If there are changes in data, such as introduction of new variable, the module appends those data without any influence to existing data and data format. If there is a new plant identified, separate file will be created automatically according to the existing data.

All the methods and algorithms were implemented using Visual Basic. Net. Six industrial scale biogas plants spread over different locations in Bryan, Germany were selected for implementing the system. Data recording module was installed in PDAs (HP iPAQ 114) and customized according to the data produces in the relevant biogas plant. An email client (Microsoft outlook 2010), ActiveSync (connecting PDA with PC), spreadsheet application (Excel 2010), and relevant data synchronizing module were installed into all client PCs (Pentium 4 or above, Windows platform) and sever side PCs (Pentium 4 or above, Windows platform).

### Results and Discussion

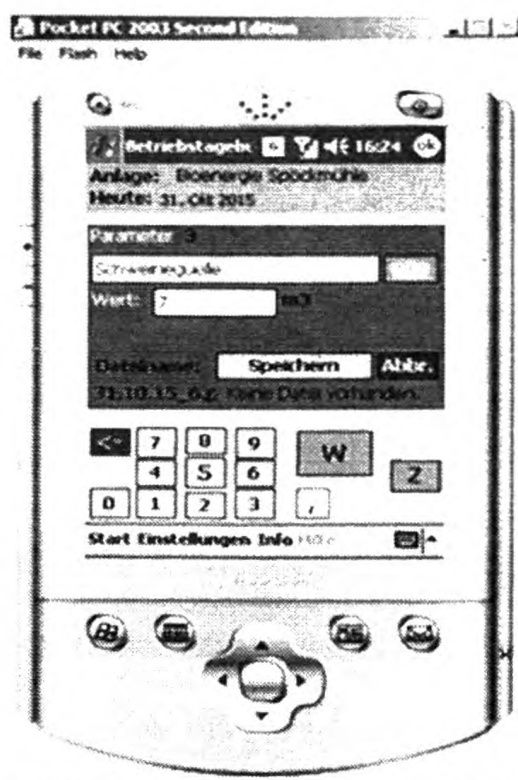
After implementing the new system we were able to first replace the data log book with a PDA (Figure 2). Before implementation of proposed method, data in the data log book were transferred into electronic format (most of the times to a spreadsheet), manually.

This process is time consuming, error prone and an inefficient method especially in the point of monitoring. With PDA, those drawbacks were eliminated. Before implementation the system data were transported manually in a USB drive to the central location. First this was time

consuming; secondly it is not cost effective due to considerable amount of transport cost. With the proposed method we were able to eliminate aforementioned drawbacks. However, the most valuable benefit was that the proposed method sends data daily to the central database, which was earlier done several times per month. Daily received data allows close monitoring of performances of biogas plants and provide necessary instructions for betterment of the biogas plants.

Because the different biogas plants use different input materials and devices from different companies, number of parameters, type of parameters and measuring units are not the same in all farms. Furthermore, even in the same biogas plant, according to the requirements input materials were changed time to time. Since the PDA system is capable of customizing the parameters and the parameter units according to the requirements of the farm, we were able to overcome this difficulty.

Earlier, the data needed for decision making, managing, accounting processes of the farm are collected manually from the data log book. With the new method, all the necessary data are automatically appended into the required format of the farm; farmers were able to save data entering time and were able to understand data easily. The advantages of the proposed method were not limited to farmers, but also to the researches, and decision makers. Proposed method provides good platform for monitoring data in relation to different biogas plants (in terms of design and technology) located in a large geographical area, as a one data collection point.



**Figure 2:** Simulator view of data entering interface of the data recording module of PDA

The proposed method can be recommended for synchronizing unstructured data as a low cost data synchronizing method. Performances of the method can be enhanced further by replacing PDA with smart phone. The built in wireless device in the smart phone can be used for sending collected data directly to the central computer as well as the local computer. Furthermore, performance of the system can be further improved by replacing spread sheet database of central database with a data storing technique that supporting unstructured data such as "data warehouse framework for unstructured biogas data" (Adikaram *et al.*, 2014).

### References

- Adikaram KKL, Hussein MA, Effenberger M and Becker T 2014. Data warehouse framework for unstructured biogas data. In: International Conference of Agricultural Engineering, Agroscope, Zurich. DOI: 10.13140/RG.2.1.3881.6489.
- Golubchik L, Khuller S, Mukherjee K and Yao Y 2013. To send or not to send: Reducing the cost of data transmission. In: Proceedings IEEE 2013. IEEE INFOCOM, Turin. DOI: 10.1109/INFOCOM.2013.6567053
- Inmon WH and Krishnan K 2011. Building the Unstructured Data Warehouse: Architecture, Analysis, and Design, Technics Publications, LLC.
- Raj P and Deka GC 2014. Handbook of Research on Cloud Infrastructures for Big Data Analytics: IGI Global.
- Sukumaran S and Sureka A 2006. Integrating Structured and Unstructured Data using Text Tagging and Annotation, Business Intelligent Journal 11(2). : 8-17.