A database-driven classification model for radio frequency interference (RFI) at MeerKAT/SKA radio telescope

Gerald Nathan Balekaki
Michelle Kuttel, Sarah Blyth & Anja Schroeder

January 9, 2017

Abstract

Radio astronomy advancements in science today have been hampered by the increasing radio frequency interference (RFI) due to technological and industrial developments. As the RFI environment continues to be polluted with electronic signals, it is critical to maintain low level of RFI and also keep the environment pristine to enable for accurate radio astronomy. The presence of RFI signals significantly reduces the sensitivity of the radio telescope, producing artifacts in the observed data. Previous work shows that there is no actual remedy for RFI problem, however there is a wide range of mitigation approaches that have been proposed ranging from offline to real-time or online processing modes. Currently, there is need for improved mitigation approaches, with more attention given to real-time and online solutions due to extreme demands in massive data handling, computational speed, and numerical precision of astronomical data. Our research seek to supply a working model of RFI database that would assist astronomers and engineers in their effort to find and classify RFI signals. The model will be able to identify any uploaded RFI occurrence by source and their nature. We also expect to introduce a means of comparison for the newly detected RFI (unknown) against a database of known RFI signals. Our Model will be follow core design principles of User-Centered Design (UCD), also evaluated for speed of data retrieval, accuracy in signal comparison, and robustness in access and big data size (astronomical data).

1 Introduction

Radio astronomers conduct studies of the celestial objects with the help of a radio telescope which uses radio frequency spectrum to better understand the phenomena of the universe. The entire electromagnetic spectrum carries information of astronomical importance, ranging from short wavelength radiations (such as x-rays or gamma rays) to long wavelength radiations (such as radio waves). However, the biggest portion of the spectrum is mainly absorbed by the Earths atmosphere [1]. Figure 1 illustrates the Earths atmosphere which is transparent to radio waves in the spectrum portion approximated from 30MHz to 300GHz or (1cm to 10m), suitable for ground-based telescope observations.

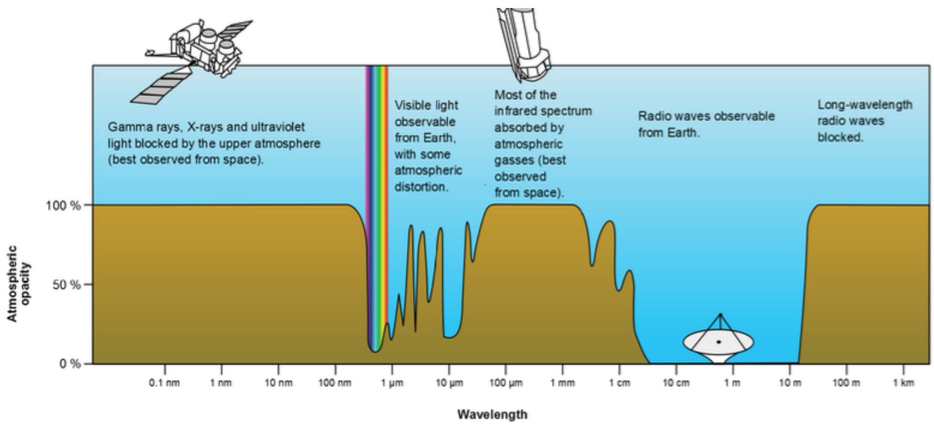

Figure 1: Earths atmospheric transmittance at different wavelength [1]

While ground-based astronomy is not affected by the Earth's atmosphere, the biggest limiting factor is mainly the man-made radio frequency interference (RFI) [2]. This kind of RFI is very severe and may lead to serious distortion of data.

MeerKAT telescope is a precursor to the Square Kilometre Array (SKA) project, and follows the Karoo Array Telescope (KAT7) that was built purposely for experimentation [3]. The MeerKAT is currently under construction in a remote site of the Karoo in the northern cape of South Africa. This site was purposely selected as a location for both telescopes as it is a reasonably radio-quiet zone suitable for scientific observations [4]. The MeerKAT will consist of 64 dishes, each comprising of 13.5m diameter dish antennas.

The Commissioning of MeerKAT is currently done in phases such that early science can be conducted with parts of the array as construction continues. The first phase (AR1) consists of 16 receptors of arrays. The subsequent phases AR2 and AR3 will consist of 32 and 16 receptors respectively [3]. The MeerKAT completion date is projected around the end of 2017 [3, 5]. The MeerKAT will be integrated in the SKA international project to build the largest and most sensitive radio telescope in the world. There will be two sites namely; one in the northern cape in South Africa (with remote stations throughout Africa) and the other in Australia.

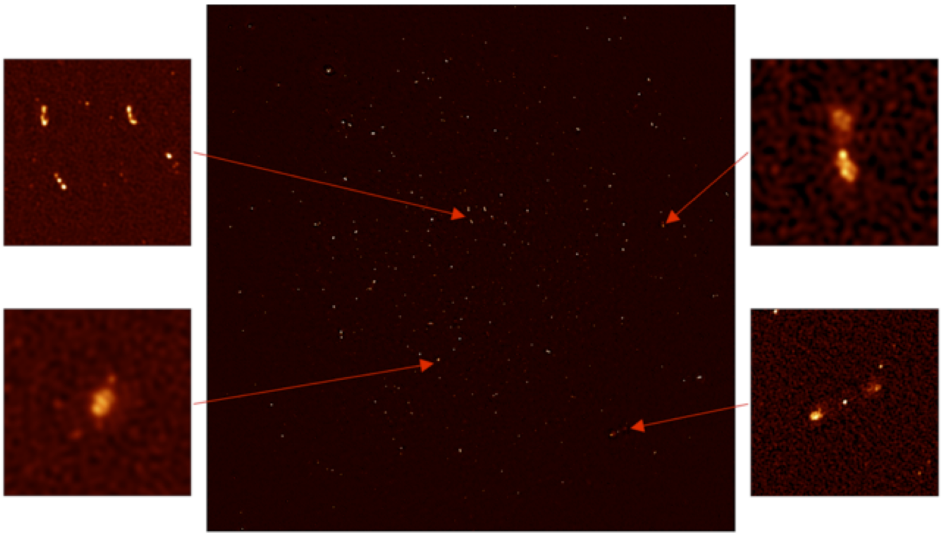

Figure 2: MeerKAT First Light Image (Photo credit: SKA South Africa) [5]

So far, the telescopes maiden milestone has been the unambiguous light image it produced on July 16th 2016. The image revealed 20 times more galaxies than previously known in the location prior to MeerKAT [5].

1.1 What RFI is in radio astronomy?

Radio frequency interference (RFI) is a recurrent problem that dates back as far as the radio astronomy inception. RFI in radio astronomy refers to any signal captured by the radio telescope which does not come from the observed target i.e. the object an astronomer observes. This may include undesired signals from other astronomical or natural objects such as the sun or any strong astronomical source [6].

Local sources of interference include things internal to telescope instruments, networking for IT systems, and general and special purpose digital processors in the observatory. Interference compliance testing, shielding, separate power circuits, minimizing nearby equipment are key steps that need to be taken to minimize this kind of interference.

External interference may arise from fixed or moving sources. Not all methods of mitigation apply to both: in fact methods that work well for fixed sources, may not work at all for moving sources, due to problems like side lobe rumble. Interference may be naturally occurring or human generated.

Examples of naturally occurring interference include: the ground, sun, other bright radio sources, and lightening. Man-made interference may come from broadcast services (eg TV, radio), voice and data communications (eg mobile telephones, two-way radio, wireless IT networks), navigation systems (eg GPS, GLONASS), radar, remote sensing, military systems, electric fences, car ignitions, and domestic appliances (eg microwave ovens) [7].

Man-made radio interference is often much (billions times) stronger [6] than the weak celestial radio emissions, thus this effect can drown-out weak signals of interest or desired signals. Modern radio telescopes today are purposely designed with highly sensitive radio receivers to detect details of the weak celestial radio emissions. The presence of RFI signals significantly reduces the sensitivity of the radio telescope, producing artifacts in the observed data[8].

The growth in RFI has been attributed mainly to increasing environmental pollution with electronic signals particularly from man-made activities [2]. Additionally, less protection of the radio band from the increased number of terrestrial radio spectrum users will lead to radio frequency (RF) congestion due to the limited spectrum [1] that may significantly cause unintentional RFI.

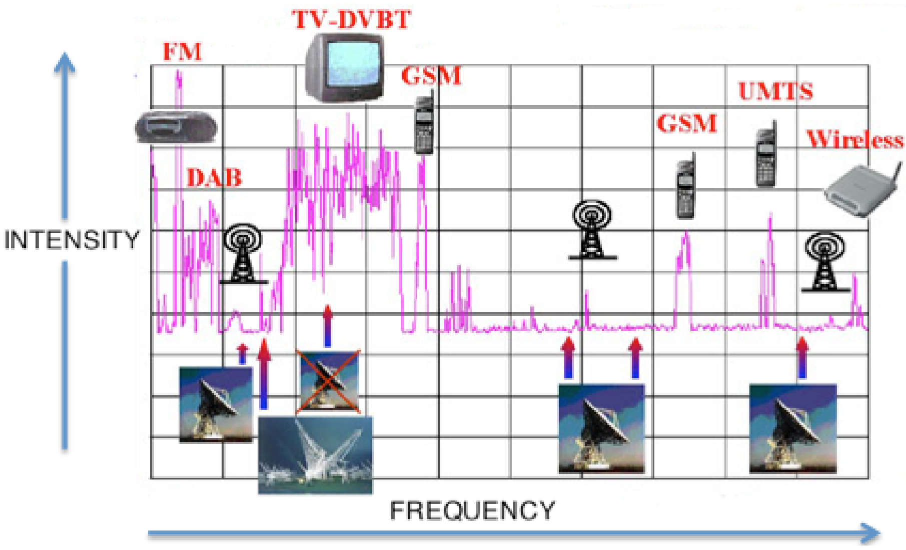

Figure 3: Potential RFI Sources near the radio astronomy bands [9]

RFI can broadly be categorized into two groups namely; narrowband RFI which mainly emanates from intentional transmissions such as television or FM radio signals while broadband RFI signals are those emitted by electric circuits or power lines, also can be referred to as unintentional transmissions [10]. It may be possible to find and shield broadband sources, more easily than narrow-band.

In the time domain, we represent a signal with RFI observed by a radio telescope as [11]:

$$x(t) = x_{sig}(t) + x_{sys}(t) + x_{RFI}(t) \tag{1}$$

Where $x_{sig}(t)$ is the contribution due to the astronomical source (desired signal), while $x_{sys}(t)$ is the system noise (undesired signal) and $x_{RFI}(t)$ is the RFI signal (undesired signal). System noise is a combination of sky background noise and receiver noise. The astronomical signal and system noise are signals with a Gaussian probability distribution with zero mean, and positive variances σ^2_{sig} and σ^2_{sys}, respectively. We note that $x_{RFI}(t)$ has a non-Gaussian probability distribution and moments are highly variable as a result of the inherent characteristics of the RFI source [11].

Figure 4 shows RFI measurements taken prior to SKA site selection, from one of the remote core candidate sites in Australia [12]. The study aimed to measure the impact of RFI on a channel over a certain period of time, such that an appropriate site with low level of interference could be selected.

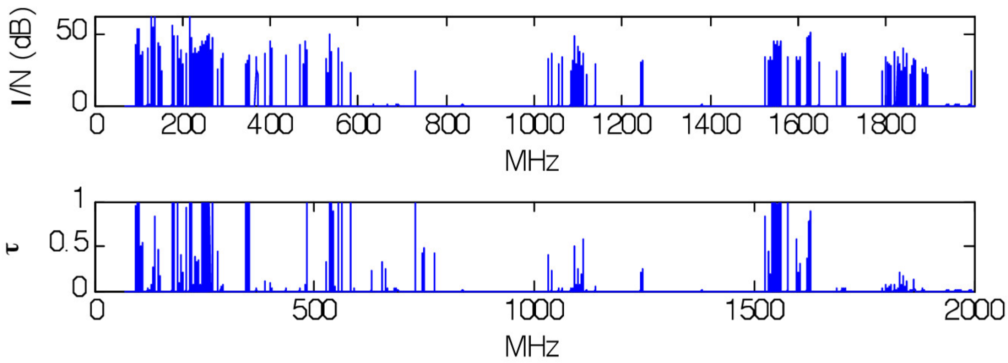

Figure 4: Interference strength and channel occupancy respectively [12]

Stronger levels of RFI over long time period and spanning the whole channel will eventually cause total loss of the channel [12]. Astronomers use occupancy statistics to determine the fraction of the time a radio telescope is available for undisturbed observations in a particular band.

1.2 Problem Statement

Radio astronomers make passive use of many parts of the spectrum legally allocated to communication and other services. As a result, many of the unwanted signals (RFI) are entirely legal and legitimate [13], but very detrimental to astronomical studies. As they are often stronger to drown out weaker astronomical signals of interest, that would result into distortion of data [2, 8, 11]. More so, it is not easy or always obvious to identify RFI, as they vary greatly in their source and nature [6]. This problem has made it difficult for radio astronomers to make appropriate choices of mitigation approach that would eventually contribute in keeping the RFI environment clean.

2 Research Motivation

Several approaches have been proposed in an attempt to effectively mitigate RFI, either by controlling, eliminating RFI sources or removing RFI from different stages of the receiver chain [8]. Previous work on RFI mitigation suggests that there is no actual remedy for RFI, however there are common techniques that have been fine-tuned to optimally work with specific telescopes as they exist in different physical, legislative and radio frequency environment [14]. For instance, with the increasing data volume available, automated techniques have been much preferred [15] to manual interventions, which are slow and cumbersome especially, in flagging large datasets.

The first approach in RFI mitigation is the prevention of RFI, this is explicitly provided in the Astronomy Geographical Act (AGA) of South Africa [4] to protect areas suited for radio astronomy. However, due to industrialization and accelerated growth in electronics, telecommunication and broadcasting, the regulations cease to be adequate as it is impossible today to allocate the complete spectrum to one group, such as the Radio Astronomers, since the spectrum has become very scarce [16].

The RFI monitoring and detection team at MeerKAT/SKA follows an pipeline approach illustrated in Figure 5 in order to mitigate RFI effectively. This involves monitoring and detection which only detect the source, but do not deal with it permanently [10]. The approach involves the use of RFI monitor, which is fitted automatically to gather data continuously from RATTY (ReAl Time Transient analYser) system. The RATTY is a development from the MeerKAT digital backend team that is based on a ROACH (Reconfigurable Open Architecture Computing Hardware) [17]. The monitor also provides access and visualization to the spectral data collected in raw formats.

Figure 5: Current mitigation approach at MeerKAT/SKA [10, 18]

In order to effectively locate the RFI source and get rid of it permanently, some studies [19, 20] have widely proposed signal characterization of RFI as a better approach. RFI signal characterization is an essential step after monitoring and detection as it becomes much easier to locate the source by either shielding, switching off the source, or dealing with the signal during data processing.

Related work on RFI mitigation majorly focussed on monitoring and detection of RFI while some recent studies [10, 18] demand for more work in the area of signal characterization, as it is noted as more effective in locating the source of interference in order to shield the interference permanently than rather to detect and deal with it temporarily. Similar work [6, 19] show that a signal can be characterized based on strength, geographical location or position of the source, polarization, direction, orientation, periodicity over time, bandwidth, frequency distributions, modulation and encoding. Some characteristics, such as strength, are easy to identify for a single source, while others, such as polarization, are more difficult to determine.

Characterization also plays a vital role in signal matching. It is easier to compare two signals that have been characterized, thus it becomes possible to identify the RFI signals through matching the similarity with the known RFI stored in one database. It is vital for users to be aware of the RFI environment around the telescope in order to be in position to make informed predictions in case of changes in environment [19].

It is important for RFI monitoring systems to avail RFI information so easily, such that the astronomers can be aware of the potential RFI as fast as possible. However, we have observed at some telescopes the monitoring system use static PDF files or offline data analysis and sometimes requires special software to provide RFI data to users. The current monitoring systems either do not provide immediate access to RFI data or required access to the RFI monitoring servers [16].

We propose the next step in the current approach at MeerKAT/SKA. Our approach will introduce a new component based on an online RFI database as shown in Figure 6. The database will be used as a basis for characterizing RFI and also monitoring the RF environment. The database will record details of all RFI occurrence, and will be used to generate relevant statistics to answer questions like; what kind of RFI occurs, when and where to enable users to intuitively better identify RFI.

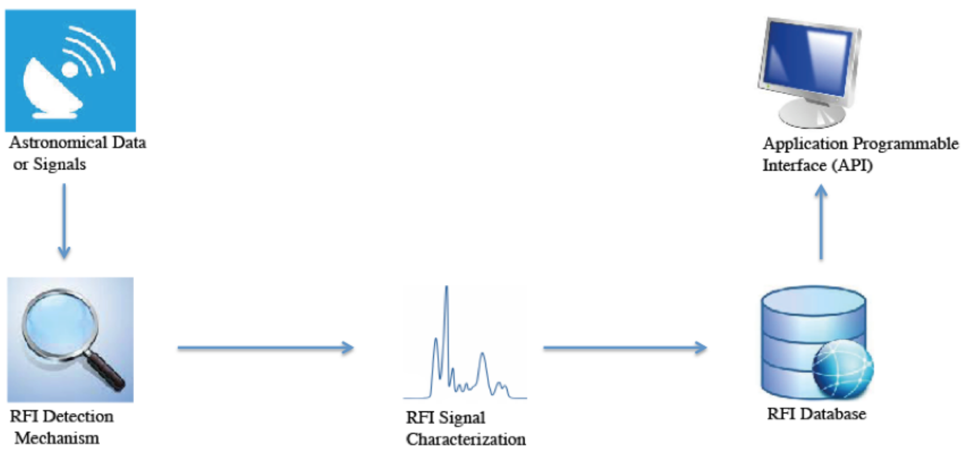

Figure 6: Proposed mitigation approach (RFI database) [10, 18, 21]

The better we can characterize RFI, the easier we can provide guidance on appropriate observational procedures and techniques to be used [11]. Our database will form part of the current RFI mitigation system at MeerKAT/SKA where RFI information is scattered and uncorrelated, making it difficult for dissemination.

User-centered design (UCD) is a broad term to describe design processes in which end-users influence how a design takes shape. It is both a broad philosophy and variety of methods. There is a spectrum of ways in which users are involved in UCD but the important concept is that users are involved one way or another. For example, some types of UCD consult users about their needs and involve them at specific times during the design process; typically during requirements gathering and usability testing.

2.1 Relational database model Vs Associative database models

One of the most commonly used database structures is the relational database model. As the name implies the relational database model has the relation at its heart, but then a whole series of rules governing keys, relationships, joins, functional dependencies, transitive dependencies, multi-valued dependencies, and modification anomalies [14 The Relational Data Model, para. 1]. Figure 3 below depicts a number of tables showing elements like table names, primary and secondary keys, field names, and relationships. While figure 3 appears to be very complex, a more simple way of looking at the relational database model is show in figure 4. This figure illustrates a simple table view, commonly seen in a spreadsheet application

2.2 Methods of Classifying RFI

2.3 Research Questions

Can we design a database that will enable real-time identification or classification of an unknown RFI signal?

2.4 Research Objectives

We aim to build an online RFI database that will be used in RFI mitigation during observations at MeerKAT/SKA.

2.4.1 Specific Objectives

- To assign a unique identification number to provide distinct information on a particular RFI signal.
- To speedup the RFI identification process to save time and money.
- To model an algorithm for signal matching to increase on the accuracy in identifying RFI signals.

3 Design

3.1 Design Goals

We aim to design a working model of a RFI database that allows for rapid characterization and identification of RFI signals. In order to achieve our ultimate design goal, we should further ensure three (3) key design aspects 1) an efficient signal storage (with the aid of data stores), 2) Signal matching by help of algorithms and 3) Speed of execution through maintaining both online and virtual platform).

3.2 Design Approach

We shall use a user-centered design (UCD) approach which involves user centered activities throughout the entire development process. The User-centered designs are recommended simply because they aid in addressing the main goal, outlines the tasks clearly, and specifies the environment that the user interacts with by the aid of softwares [22]. In this project, we identify intended users as scientists, researchers, engineers in the field of radio astronomy as our key stakeholders.

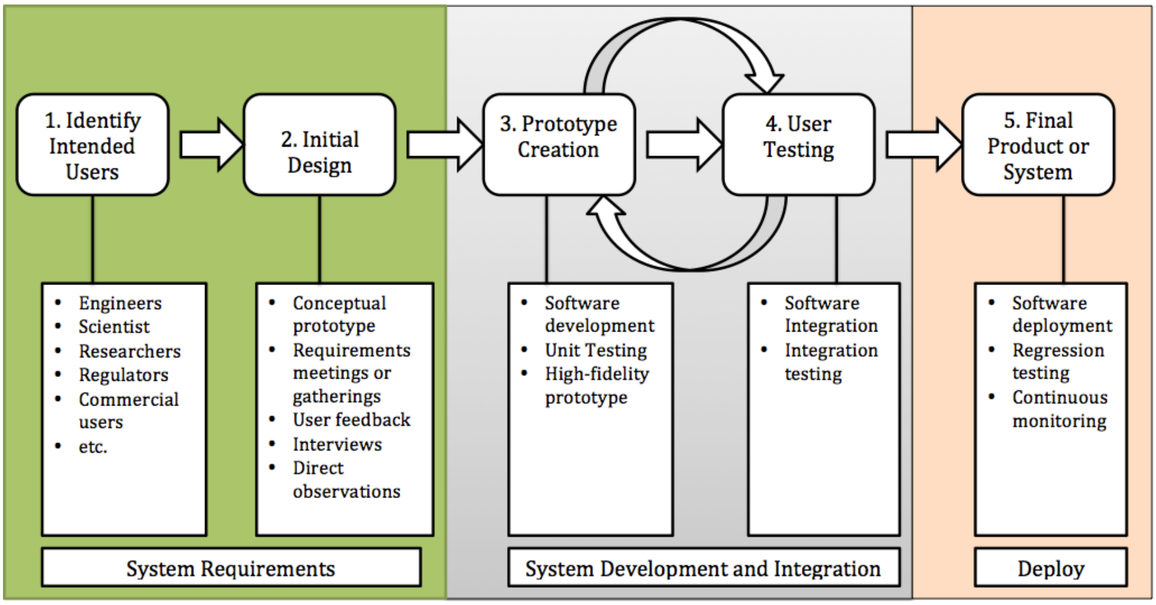

Figure 7: Steps in the user-centered design process

Figure 7, illustrates the steps involved in this process as: (1) determine the intended user of the software; (2) interviewing users to inform the initial design; (3) creation of the prototype; (4) a multi-stage process of user prototype testing; and (5) creation of a final product or system. The user-centered design process integrates feedback from the user at every stage in the development of the software.

3.2.1 Identify Intended Users

The first step in the user-centered design process is to identify the user of the software. The intended users of this specific system are basically astronomers, engineers at MeerKAT/SKA. In Table 1, we present a list of some of the potential RFI (culprits) detected at MeerKAT/SKA telescope. The table broadly classify each RFI signal basing on the type of transmission as intentional or unintentional [16].

Table 1, further divides the RFI signals observed, into smaller groups basing on where the are emitted from (source) e.g. ground-based, space-based, aircraft-bases etc. We also listed key parameters that are used in identifying uniquely a particular RFI signal. We shall study, match and analyze each parameter in a given group. The information forms part of our requirements gathering task required in this project as shown in our design guidelines in Figure 7.

RFI Type	RFI Groups	RFI Parameters
Intentional	Ground-based/ Fixed Transmitters: (E.g Radio, TV, GSM, Data Links, Fixed Position, High site, High Power, Omni or directional antennas	Frequency Bandwidth (analogue/digital) Intensity Frequency time Occupancy (short/long term) GPS Location
	Space-based Transmitters: (E.g GPS, Galsnost, Orbincom, DTT, Geostation or not, High Power, Large Footprint)	
	Aircraft Transmitters: (E.g VHF Comms, Transponders, DME, Radar)	
Unintentional	Short Range Transmitters: (E.g Wifi, DECT, Bluetooth, GSM, Handset, VHF Handset)	
	Controller and Processors Transmitters: (E.g Micro Processors, Micro Controller, RFPGA, GPU)	Base Frequency Harmonics Frequencies/Spacing
	LCD Transmitters: (E.g PCs, Cameras, Smartphone, Smart watches)	Bandwidth Shape
	Power Supplies Transmitters: (E.g Charges, UPS, PC UPS)	
	Short-Duration Transmitters/Transient: (E.g Electrical switching, Realys, Pulses)	Time based

Table 1: Source: RFI Working Group, SKA Offices, Cape Town, RSA [16]

.

3.2.2 Initial Design

The initial design of our system will be based on the feedback from our intended users through interviews, direct observation (HartRao study trip), focus group meetings (Monthly RFI meetings at SKA offices) and also technical guidance from the RFI Manager at SKA (Mr. Merwe Carel). Throughout this process, individuals will identify unique needs and desired capabilities of the software to be developed. Based on our initial interactions with the intended users, we have come up with a conceptual prototype in Figure 8.

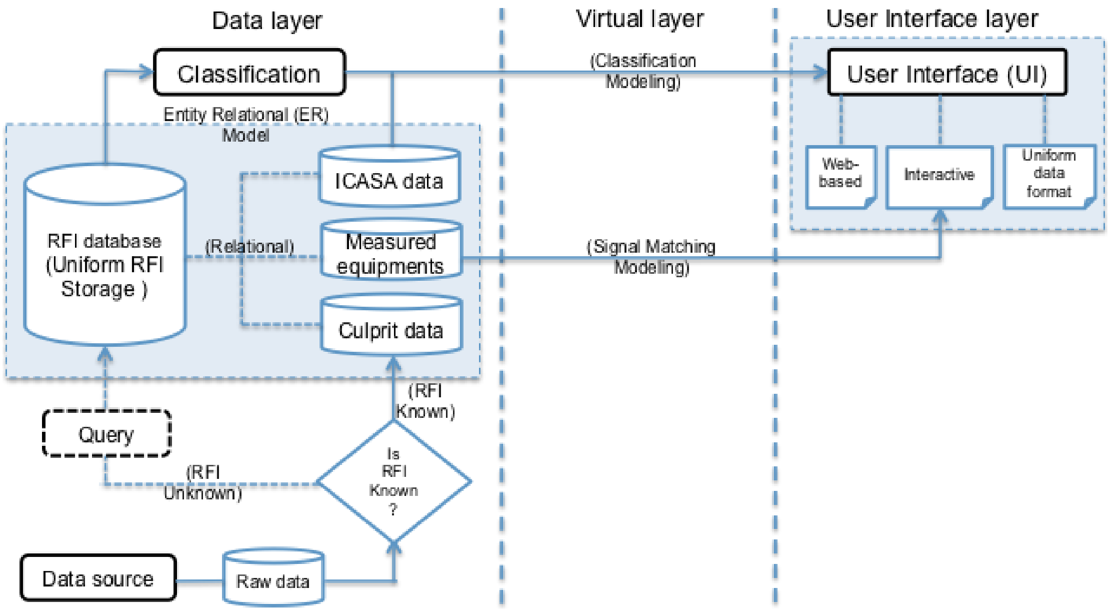

Figure 8: Conceptual Prototye

We expect our database to provide an interactive platform, with dynamic and well formatted interface. We also plan to run our database on MySQL engine with the aid of the LAMP functionality (Linux, Apache, MySQL, PHP/Perl/Python) [23], such that we can maintain the current formats of the RFI scans from our RFI Monitor at MeerKAT. We hope to maintain two storage platforms namely; at the local device on the network (at Karoo site), and also on cloud for users Offsite (SKA offices, Cape Town, Observatories, Universities etc.) We plan to have a record of all detected RFI in one database, and in cases of unknown RFI, we then characterize and also match the signal together with our predefined parameters in the database. With this, the RFI signal can be easily identified thus giving guidance before and during the observations.

3.3 Evaluation

Our model will be evaluated based on our key design aspects by measuring; 1) speed of data retrieval 2) accuracy in signal classification and comparison (spectra data), and 3) robustness in access and data size (big astronomical data). We plan to measure the speed in terms of response time in seconds, and accuracy by comparing ratios of known to unknown RFI signals. While robustness will be tested through generating large data sets of varying data formats, simultaneously into a single database model.

3.4 Expected Outcomes

We expect to provide a working model of RFI database that would assist astronomers and engineers in their effort to find and classify RFI signals. The model should be in position to identify any RFI occurrence by source and their nature. We also expect to introduce a means of comparison for the newly detected RFI (unknown) against a database of known RFI signals. With these expectations under consideration, we hope to keep the RFI environment of the radio telescope crystal clean.

3.5 Ethnical issues

We note that our research work does not pose any unethical threat or issue but rather adhere faithfully to the norms of our profession in the society. We promise to maintain integrity, also address any ethical issue that might arise and make ethically sound decisions while dealing with them.

3.6 Research Timeline

Planned Steps/ Activities	**Deadline (By Month & Year)**
Submission of the proposal	September, 2016
Proposal Presentation	October, 2016
Design of a research plan (UCD)	October, 2016
Requirements gathering	November, 2017
Initial Design/ Conceptual prototype	December, 2016
Literature review	March, 2017
Prototype creation	April, 2017
Software development & Integration	October, 2017
System Testing & Evaluation	December, 2017
Report on the final system	March, 2018
Presentation of final research results (Conference)	September, 2018
Thesis Writing (First Draft)	December, 2019
Second Draft (Peer reviewed)	February, 2019
Final Write-up	March, 2019
Book Binding & Submission	April, 2019

3.7 Milestones

Major Milestones	Oct,16	Nov, 2016	Dec, 2016	Feb, 2017	Mar, 2017	Apr, 2017	Oct, 2017	Nov, 2017	Dec, 2017	Apr, 2018	Dec, 2018
Complete Proposal writing	Format document										
Cont'd requirement gathering	Interviews, SKA meetings, discussions			Review the user requirements				Conform with initial user requirements			
Initial Design		Integrate feedback, develop conceptual prototype									
Setup Development				Develope prototype							
Software dev't & integration				Coding & unit testing							
Test system & Evaluate							Validate the system				
Final System								Acceptance testing			
System Goes Online											Launch online

References

[1] B. Weeden, "Radio Frequency Spectrum, Interference and Satellites Fact Sheet," *Secure World Foundation*, Jun. 2013.

[2] J. G. Porko, "Radio frequency interference in radio astronomy," Master's thesis, Aalto University School of Electrical Engineering, Finland, 2011.

[3] MeerKAT/SKA, "MeerKAT AR1 Array Release 1: Information Sheet," *Square Kilometre Array (SKA South Africa)*, Jul. 2016.

[4] Department of Science & Technology (DST), "Astronomy Geographic Advantage Act," *National Gazettes, No 37397. Republic of South Africa*, 2007.

[5] MeerKAT/SKA, "MeerKAT radio telescope. Internet: www.ska.ac.za/gallery/meerkat/," Sep. 14, 2016 [Sep. 16, 2016].

[6] R. D. Ekers and J. F. Bell, "Radio Frequency Interference," in *arXiv:astro-ph/0002515v1*, Feb. 2000.

[7] M. Goris, "Categories of Radio-Frequency Interference," *National Foundation for Radio Astronomy*, Feb. 1998.

[8] J. M. Ford and K. D. Buch, "RFI mitigation techniques in radio astronomy," in *Proceedings IEEE Geoscience and Remote Sensing Symposium*, p. 231234, Jul. 2014.

[9] G. Bianchi, "Interference Monitoring. Internet: www.med.ira.inaf.it/Interferenze_page_EN.htm," Jan. 14, 2013 [Sep. 16, 2016].

[10] P. Hillebrand, "Detection and Visualisation of Radio Frequency Interference," *University of Cape Town, South Africa*, Oct. 2014.

[11] P. A. Fridman and W. A. Baan, "RFI mitigation methods in radio astronomy," in *Astronomy and Astrophysics*, vol. 378, pp. 327–344, Aug 2001.

[12] SKA Expert Panel, "A report on the strengths and weaknesses of the currrent radio frequency interference environment as measured at the SKA candidate sites," *Square Kilometre Array*, Nov. 2011.

[13] Independent Communications Authority of South Africa (ICASA), "National Radio Frequency Plan 2013," *National Gazettes, No 36336. Republic of South Africa*, 2013.

[14] W. A. Baan, "RFI mitigation in radio astronomy," in *Proceedings General Assembly and Scientific Symposium*, Aug. 2011.

[15] R. Millenaar, "System Approach to RFI Mitigation for the SKA," in *Proceedings of Science*, Mar. 2010.

[16] C. Merwe, "Culprit and victim management RFI environment for a radio astronomy site," Master's thesis, Stellenbosch University, South Africa, 2012.

[17] A. Botha, "Development of a Real-Time Transient Analyser for the SKA," Master's thesis, Stellenbosch University, South Africa, 2013.

[18] C. Schollar, "RFI Monitoring for the MeerKAT Radio Telescope," Master's thesis, University of Cape Town, South Africa, 2015.

[19] R. Oliva and E. Daganzo et al, "SMOS Radio Frequency Interference Scenario: Status and Actions Taken to Improve the RFI Environment in the 14001427-MHz Passive Band," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, pp. 1427 – 1439, May. 2012.

[20] K. Player and T. Stumpf et al, "Characterization and Mitigation of RFI Signals in Radar Depth Sounder Data of Greenland Ice Sheet," in *IEEE Transactions on Electromagnetic Compatibility*, vol. 55, pp. 1060 – 1067, Jun. 2013.

[21] A. Jessner, "Committee on Radio Astronomy Frequencies," *European Science Foundation*, Jul. 2010.

[22] R. R. Hall, "Prototyping for usability," in *Proceedings International Journal of Human-Computer Studies*, vol. 55, pp. 485–501, 2001.

[23] ORACLE Inc., "Top 10 Reasons to Choose MySQL for Web-based Applications," in *Proceedings Oracle and affiliates*, Aug. 2011.