Towards Data-Driven Policies in Spectrum Management

Karyn Doke, †Ali Abedi, ‡Max Hollingsworth, Mariya Zheleva, †Anant Sahai, ‡Dirk Grunwald, §Keith Gremban
Department of Computer Science, University at Albany, {kdoke, mzheleva}@albany.edu
† *Department of Electrical Engineering and Computer Sciences, UC Berkeley*, {abedi, sahai}@berkeley.edu
‡*Department of Computer Science, University of Colorado Boulder*, {max.hollingsworth, dirk.grunwald}@colorado.edu
§*Ann and H.J. Smead Aerospace Engineering Sciences, University of Colorado Boulder*, {keith.gremban}@colorado.edu

***Abstract*—This position paper explores the gap between the current state-of-the-art in spectrum management and the objective of data driven spectrum policy. We explore four issues underlying successful data-driven policy: data requirements to support policy decisions; data acquisition and storage; robust, extensible metadata; and tools for analysis and visualization. For each issue, we discuss the state-of-the-art and describe the ultimate objective. We conclude the paper with a call for action to the spectrum community and list a number of efforts that should be undertaken to support true data-driven spectrum policy.**

I. Introduction

Radio frequency (RF) spectrum is an increasingly scarce resource, owning to the popularity of mobile and wireless broadband and the increasing need of wireless spectrum in other domains, such as radio astronomy and remote sensing [1]. In 2016 Weldon [2] projected that utilized bandwidth would increase by 10x between 2010 and 2024; that network access would shift from 97% wired to 97% wireless; and that the number of wireless household devices would increase by 10x.

To make radio frequency (RF) spectrum available for new services, regulators are forced to re-allocate spectrum from existing services (incumbents), or develop mechanisms to share spectrum between incumbents and new entries. This process was formalized in the U.S. with a 2010 Presidential Memorandum [3] which tasked the National Telecommunications and Information Administration (NTIA) to identify 500 MHz of government spectrum that could be re-allocated. Some examples of spectrum reallocation include:

- The 600 MHz Transition reallocated 70 MHz of spectrum formally allocated for broadcast television to wireless services [4];
- The Advanced Wireless Services 3 (AWS-3) transition reallocated the 1695-1710 MHz, 1755-1780 MHz, and 2155-2180 MHz bands from federal to commercial use. Some federal systems were unable to relocate and will share spectrum with new entries [5];
- The Citizens Broadband Radio Service (CBRS) band, 3550-3700 MHz was reallocated for non-federal use via a complex sharing mechanism with DoD radars [6].

When making spectrum management decisions including reallocation, assignments, and coexistence analysis and techniques, regulators must consider a number of technical questions including, but not limited to:

(1) What is the incumbent occupancy in key geographic areas? Policy examples include TV White Space (TVWS) [7] and CBRS [6].
(2) What are the propagation characteristics of the band in consideration? Relevant policy examples include the effect of propagation models on TVWS density [8].
(3) What are the electromagnetic compatibility characteristics of the involved and neighboring systems? Most of these policy decisions concern interference based on the signal strength or transmission mask of adjacent channels as in the role of Dedicated Short Range Communications (DSRC) in the 5.9Ghz [9]. Issues such as Ligado/LightSquared involve receiver specifications, but knowledge of the existing interference in those bands could influence policy decisions [10].
(4) What is the potential for *aggregate interference* from a large numbers of new devices? Relevant examples of the effects of aggregate interference include: Industrial, Scientific, and Medical (ISM) bands [11], LED interference [12] and remote sensing [13].
(5) How can the effectiveness or efficiency of a policy be evaluated? Are the policy decisions meeting their stated goals? This question has started to be explored for the case of CBRS, but typical efforts to date have involved counting devices as opposed to measuring occupancy [14].

The FCC and NTIA have long sought a more *data driven* approach to spectrum management and allocation policies. In the Spring of 2023, under the auspices of the Wireless Spectrum R&D Interagency Working Group (WSRD), and hosted by the National Institute for Standards and Technology (NIST) a workshop was held titled ”Making Data Available for National Spectrum Management.” [15]. At the workshop, several points were made about data driven spectrum policy:

- The goal is data driven policy, but how to achieve this?
- The community needs meaningful guidance for spectrum data collection.

The question naturally arises that, *given the number of organizations (academic, government, industry) making spectrum measurements of various types, why is the resulting data not* *sufficient to drive data driven policy*? In this position paper, we explore this question by examining the requirements for spectrum measurements, collection, distribution and cataloging for data driven policy decisions. Section II discusses the requirements for data driven policy. Section III surveys existing data collection systems. Section IV explores the need for good metadata. Section V explores the requirements for analysis and visualization. The paper concludes with a Call to Action to the spectrum community with recommendations to enable data driven spectrum policy.

II. Requirements for Data Driven Policy

The point of policy is to provide the stability required to allow *future* investments in using the relevant spectrum in a manner that delivers useful benefits for people. By contrast, actual data is *always* about the past or the present — we never have direct access to the future. Consequently, data-collection impacts policy making in two distinct ways:

- Via implicit/explicit forecasts of conditions in the relevant future. This is the only path by which actually collected data can be useful for policy making.
- As a capability that one can count on as a part of the policy itself. Reliable data in the future allows *closed-loop* policies that adapt to local circumstances.

Specific parameters of any policy that made sense with a particular view of the state of the world (including uncertainty estimates) can be less sensible for the more refined view that fresh data can enable. Being able to adapt those parameters in "closed-loop" is a powerful tool in improving the efficient utilization of spectrum. However, a successful adaptive policy has at least six prerequisites:

(a) Data collection capabilities that can provide us with policy-relevant data across time/space/frequency as relevant to the policy parameters being adapted.
(b) We must have a practical level of trust in the data coming from these capabilities.
(c) The capability to appropriately disseminate the relevant updated policy parameters to spectrum-using agents in an automatic and timely manner.
(d) Spectrum-using agents must be able to actually adapt their usage characteristics to the updated parameters.
(e) We must have a practical level of trust that spectrum-using agents can/will adapt to updated policy parameters.
(f) Those developing/deploying spectrum-using agents must be able to forecast the range of policy parameters expected in the useful lifetime of spectrum-using agents in order to make investment decisions rationally.

The last point (f) above is simultaneously banal — this is a characteristic of any real-world investment scenario involving any resource, not just spectrum — and vital since such stability is the whole point of policy in the first place. Without sufficient stability, investments do not happen. A quantitative sense of historic trends, variability, and correlations can help manage risks, and this requires well-instrumented, dense, and long-term longitudinal data. **This value of good data is one of the underappreciated roles for spectrum monitoring.** Good hierarchical visualizations that permit rich multimodal exploration and interaction by diverse human decision-makers and stakeholders are also vital in both taming the fear of "unknown unknowns" that are the bane of risk-takers, as well as refining the sense of relevant "known unknowns" that need to be quantified — possibly by commissioning additional measurements or collecting additional types of data.

Points (a), (c), and (d) are within the mainstream of engineering thinking for dynamic spectrum, although a lot more needs to be understood. However, to actually have data driven policy, the trust dimensions represented by (b) and (e) are critically important and much more engineering research and even basic problem formulation is needed. Understanding the tradeoffs that underlie being able to determine what "practical level of trust" should mean is one critically understudied dimension here, as is understanding the different aspects of the real world that can potentially undermine trust.

A. Spectrum Data

Each technical question described in §I needs different data:
Band Power: Incumbent occupancy measurements require at least power or carrier sensing and can record specific frequencies occupied, the duration of those events and *e.g.* detected signal power. This is used for cases (1), (3), (4) and (5) described in §I.
Adjacent Band Power: Band occupancy measures the use of a band, but certain policy decisions depend on the activity in adjacent bands, including aggregate interference from those bands (*e.g.* (3) and (4) from §I)
Data Measures: In certain cases, it may be useful to record IQ samples or band-specific derived products - for example, when determining the need for additional bands for specific services (LTE, WiFi, *etc*), it can be useful to record the *data* utilization rather than just transmission power. A 20MHz LTE signal will have some degree of constant, detectable power but may be conveying very little data. This is used for case (5) described in §I.
Propagation characteristics: Signals are affected by transmitter and antenna characteristics, elevation above ground and environmental effects. Predicted interference relies on more detailed and careful measurement that matches the policy goals such as (2) and (4) in §I.

Although IQ samples are the standard from which other data can be derived, IQ is expensive to collect and store and not needed for many policy decisions. Surprisingly, most existing data collection systems focus on recording IQ samples.

III. Existing Data Collection Systems

Spectrum monitoring has been performed by governments [16], researchers [17] and companies [18]; most monitoring systems use a single sensor at each location and those locations are widely separated. Such systems can be used to determine how spectrum is used in order to justify or argue spectrum policy allocations, to measure increases in the noise floor, or to indicate that interference has occurred.

In the realm of large-scale spectrum monitoring, numerous studies advocate for distributed data collection using affordable, software-defined radio (SDR) sensors such as SpecSense [19] and the systems proposed in [20], [21]. While SpecSense emphasizes sensor selection optimization to minimize deployment costs, both rely on centralized cloud processing for efficient data aggregation and analysis. SpecSense focuses on on-demand spectrum occupancy queries, utilizing sensor augmentation and interpolation techniques to fill data gaps. The system in [20] takes a broader approach, leveraging a secure network for collaborative sensing and data sharing, ensuring accuracy and coverage across the wideband spectrum; [21] explores low-cost SDRs that are directly connected to a smartphone for real-time spectrum monitoring suggesting the potential for on-device processing and analysis.

ElectroSense [22] is a crowd-sourced spectrum monitoring network designed to oversee the spectrum on a large scale through the integration of low-cost spectrum sensing nodes. This innovative platform not only democratizes access to crucial spectrum information but also ensures that the monitoring process is cost-effective and scalable. By deploying a network of affordable spectrum sensing nodes, ElectroSense empowers users to contribute to a collective effort in monitoring and understanding the radio frequency spectrum. This collaborative approach facilitates comprehensive coverage and also promotes real-time data acquisition, allowing for a more dynamic and responsive spectrum monitoring ecosystem. ElectroSense has been used in a variety of studies, such as real-time wireless technology classification [23].

ElectroSense supports both Power Spectrum Density (PSD) and IQ data pipelines. For PSD data, the Fast Fourier Transform (FFT) is computed. Subsequently, the derived PDF data is compressed and permanently stored in a dedicated database for future processing. Unlike PSD data, ElectroSense does not permanently store IQ data due to its inherently large size.

KiwiSDR offers a unique and flexible approach to radio signal processing. The architecture of KiwiSDR is centered around a low-cost, credit-card-sized BeagleBone single-board computer, that serves as the processing engine. Connected to this board is a high-performance wideband RF frontend that operates in the HF band (i.e., up to 30 MHz). The distributed nature of KiwiSDR allows multiple users to access the system simultaneously, each with their own independent tuning and demodulation settings. The data pipeline of KiwiSDR is designed to efficiently handle the streaming and processing of RF data, ensuring low-latency access for users.

Some automated spectrum interference detection systems have been proposed – one of the oldest involved comparing measured emissions to a database of licenses for land-mobile radio [24] and more recent versions [25], [26] have a similar design. Companies [27] have deployed similar systems – CarrierIQ's system collects "key performance indicators" that can be used to detect interference.

An interesting and vital question is why, despite these numerous spectrum monitoring systems, data-driven policy has not been realized?

The primary reason is that these are developed as "general platform", whereas policy-making (and other applications) require greater specificity in measurement capabilities. Let us take a look at one example candidate aspect of policy innovation to understand this better. Several policy decisions differentiate between *indoor* and *outdoor* transmission limits. For example, in the CBRS, transmission power limits vary for indoor and outdoor installations [28]. Similarly, in the case of WiFi 6E, deployment is restricted to indoor environments to prevent potential interference with outdoor fixed microwave links and the FCC has already revisited WiFi transmission levels in the 5GHz U-NNI-1, U-NNI-2A and U-NNI-4 bands similar to the Low-Power Indoor limits in WiFi 6E. Imposing distinct power thresholds is intended to reduce interference, ensuring the coexistence of different applications within the shared spectrum.

As a result, accurate detection of indoor versus outdoor environments becomes pivotal for these applications. This enables the implementation of adaptive power control strategies, facilitating harmonious spectrum sharing. Abedi et al. [29] have introduced a technique for automatic indoor-outdoor detection using ADS-B signals from airplanes. Their findings indicate that the conventional practice of adjusting transmission configurations based on indoor or outdoor categorization may lack effectiveness in various environments. To our knowledge, no existing spectrum measurement system captures such indoor/outdoor distinctions and could provide the data required to tune/adjust policy in this direction.

IV. Metadata is Essential

Spectrum data collection efforts thus-far have been largely decoupled from the analytics task the data will support. At the same time, spectrum analytics rely on the quality of the underlying sensor properties, configuration, behavior and environment. Coupling these properties with the fidelity of spectrum data and the corresponding spectrum analytics outcomes must also be handled by metadata [30]. Those factors can vary between different sensors and even similar sensors in multiple vantage points. The interpretation of the underlying data, and its suitability for policy decisions, depends on these factors and thus they must be documented. Subsequently, the notion of *spectrum metadata* has emerged, resulting in several proposed metadata standards including VITA49 [31], CHDR [32] and, most recently, SigMF [33], which underpins the IEEE 802.22.3 SCOS [34]. Fig. 1 summarizes data properties that can be currently captured with existing metadata standards (in red) and emerging data characteristics (in black). Standardizing metadata also involves entailing specialists in at least GIS, radio science and mobility. For example the **FixedRoute** data might be represented as a series of GIS points or a `LineString` that captures an entire path or also include altitude and velocity. We believe it is essential to provide prescriptive practices documenting what metadata is necessary for different spectrum recording activities to provide that trust to data consumers.

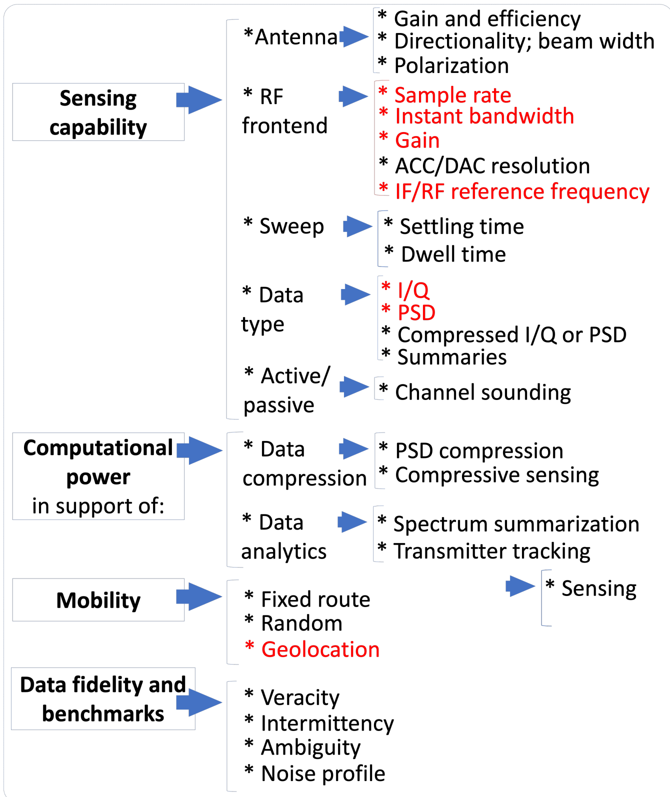

Fig. 1: Metadata capabilities vs. requirements. Highlighted in red are sensor capabilities that can currently be captured by existing metadata standards. Expanded in black are additional capabilities that will ensure data is sufficiently coupled with the collection approach and can provide trustworthy analytics.

VITA49 is designed as a "digital IF" to describe the steps and components a signal traverses during transmission or reception and is not directly applicable for archival data. The standard includes ∼22 fields, some of which are highlighted in red in Fig. 1. The fields describe the configuration of the analog RF-digital chain, and some environmental factors (temperature, GPS coordinates) but not other factors such as antenna type, antenna height, *etc*. VITA49 is a packet based protocol and the context data is embedded within the data making it difficult to search for the metadata fields.

CHDR [32] was developed by Ettus Research and is currently embedded in all USRP B2XX products. It is designed as an RFNoC transport protocol and contains fields for "user-defined" metadata, which are not specified by the CHDR standard itself, leaving them open for custom implementation.

SigMF is explicitly designed for recorded data samples, making it particularly relevant for data-driven policy. It features a prescriptive data format with fields highlighted in red in Fig. 1. The metadata is stored in hierarchical JSON and explicitly designed to support extensions. The standard has a core data schema and several extensions including those provided by NTIA [35] that define environmental, sensor, waveform and processing algorithms among others.

Metadata Design: We believe that having a metadata format separate from recorded data, as in SigMF, is important for RF archives because a key activity of an archive is discovery and retrieval of different datasets. Embedded metadata within an IQ stream makes it difficult to search because the full dataset must be scanned; while the embedded metadata can be stored in a separate form, neither VITA49 or CHDR define such a form. Yet, we believe that SigMF requires specific extensions –it is important that multiple data formats can be described by the metadata and that it is important to describe "derived products" and the provenance of those products. For example, a spectrum occupancy measurement of a time-granularity average and peak power measurements or a frequency-based occupancy measurement using FFT's would be derived from previously recorded IQ samples. Because those different products may reside in different spectrum repositories, mechanisms such as Digital Object Identifiers (DOI's) [36] to link one data product to another are essential metadata extensions. Similarly, while most spectrum data can be represented as time-series vectors, numerous concrete representations exist that more easily facilitate cloud storage.

Query Methods: Metadata is only as useful as the mechanisms to query that metadata and the underlying data. Querying RF archives involves both a mechanism and a query language. While REST API's are the most common query mechanism, they are poorly adapted to extensible schema; a typical REST API uses a set of verbs (PUT, GET) over a set of nouns (user name, location). We believe resolving policy questions using spectrum databases will typically require more complex queries such as "*the time of peak power measurements from PSD's collected in 700Mhz-750Mhz in Ohio collected using omnidirectional antennnas outdoors no more than 10 meters above ground*". Because REST API's operate on a primary noun, multiple REST queries need to be performed to retrieve complex data relationships. For example, a query to retrieve datasets using a specific antenna type may return a long list of identifiers, each of which may need to be queried to determine if that record was then in a frequency of interest. GraphQL is designed to reduce the number of those queries by combining queries across aspects of a recorded metadata and possibly for time-series data.

There are several query languages for time-series data that focus on vectors of values and the ability to operate over ranges of that data in a query. For example PromQL [37] can query data at an instant in time or a range of time values across multiple time-series datasets; however, it is designed for simple scalar values although extensions to query histograms have been added. We know of no suitable standard for locating RF datasets that match specific sensor (*e.g.* antenna pattern) and dataset (*e.g.* carrier frequency).

Data Discovery Methods: Regulatory bodies need to understand the provenance of data and be able to cite specific datasets and the relationships of different datasets to one another. Even with standardized data formats and metadata descriptions, the problem of searching for and referencing specific datasets needs study.

The DOI system [36] is designed to relate unique digital object identifiers to specific objects coupled with metadata to

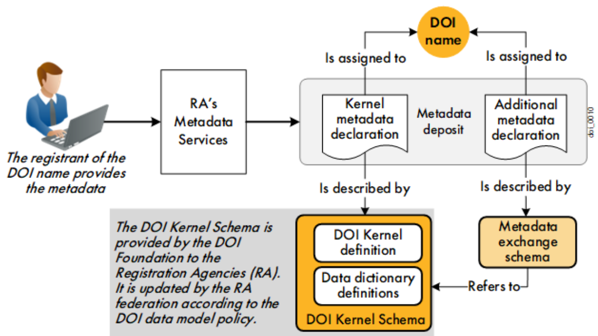

Fig. 2: Diagram from [36] showing the process of registering a DOI with a associated reference and metadata.

describe the object. DOI was originally created for entertainment artifacts, but includes different registration authorities that maintain specific DOI areas and the metadata for them. DOI is an existing standard used by government agencies – for example, the NSF uses the ORCID registration authority for researcher identities and to relate identify publications specified by a DOI to those researchers. The DOI system has been used to identify datasets – one such registration authority is https://datacite.org/. As shown in Figure 2, DOI is similar to DNS by providing unique names for a dataset and allowing the location of that data to move if needed. The DOI system does not record the data itself nor does it fully describe that data - it simply provides the references to it and methods to locate it. We believe that the spectrum science community should adopt a DOI namespace to refer to RF datasets. The SigMF metadata format already includes DOI fields for data and metadata in the `global` namespace.

Uniquely identifying a dataset leaves open the problem of data discovery. Despite efforts at developing singular data archives, the parallel history of documents in the WWW has shown that multiple repositories (websites) and search-based methods (*e.g.* google search) provide advantages. Developing a search method for RF datasets should be studied. IQEngine [38] demonstrates a newer approach by using a Large Language Model to ingest the SigMF metadata descriptions and then using free-text queries to locate relevant datasets. Once a standard metadata and DOI framework is adopted, research will be needed on effective data discovery.

V. Analytics and Visualization

Several spectrum monitoring solutions and web-based toolkits for analyzing, processing, and visualizing data have been designed and proposed in literature.

IQEngine is a web-based toolkit for sharing and analyzing RF IQ recordings. The web interface allows for easy access to view and analyze IQ recordings and allows users to manually annotate them to highlight known spectrum activity. Recordings can be visualized as spectrograms, in time and frequency domain, or IQ timeseries. The interface includes tools to adjust the display colormap, FFT size, window, filters, and resolution.

GR-Bokehui and Gqrx are open-source graphical toolkits that can visualize IQ traces in a waterfall plot. Electrosense [22] allows interactive sensor selection and varying frequency resolution. Historical data from the sensors can be viewed and live spectrum updates can also be displayed. A similar approach was taken by the Microsoft Spectrum Observatory [39], which used higher-end RFEye and USRP sensors and provided a web-UI for data visualization. Basic analytics, such as fraction of band occupancy based on power thresholding, were also available through this platform. Both the Microsoft Spectrum Observatory and Electrosense were decommissioned since their inception. All these general tools take in IQ data and not policy specific datatypes.

Some vertically integrated systems have been studied. The authors in [40] propose a database-assisted spectrum sharing system that takes in information on license holders, their rights, and policy constraints. It uses algorithms and a variety of data to make spectrum sharing decisions, but provides no visualization tools. Cosmic-CoNN [41] is an interactive image segmentation and visualization framework for cosmic rays and demonstrates the advantage of domain-specific RF analysis systems. It allows identifying, inspecting, and editing of tiny objects in large multimegapixel high-dynamic-range images. This work attempts to create an end-to-end system and displays the identified object in a visualization tool.

Collection and basic visualization of raw spectrum data only goes part of the way towards comprehensive support of data-driven policy. In addition, required are spectrum analytics toolkits, which can automatically and autonomously mine spectrum data for insights that inform policy. Such toolkits should be embedded with the visualization engines to allow policy-makers to toggle between insights and the underlying data. For example, a policy-maker trying to discern aggregate interference would require spectrum analytic tasks such as transmitter localization. A policy-maker trying to reallocate spectrum may need to assess who is using the spectrum (e.g. signal classification), and what resources are being used. Data driven policy decisions require specific analytic tasks. While such tasks have been researched [19], [21], [23], [42], existing visualization tools do not support such tasks required to inform policy. Furthermore, existing analytics methods are often developed for offline mining and have not been tested on longitudinal spectrum data. The latter, however, is critical to enable autonomous annotation of spectrum activity, and create spectrum utilization stories from the raw data.

While these existing systems provide the tools to visualize spectrum data (i.e. observe the distribution of radio signals in terms of frequency and time), data driven policy decisions require specific data analytic tasks that are not supported by them. Analytic and visualization tools should be developed in tandem and support insights to spectrum utilization, analytics required to make policy decisions, and trust in data products.

VI. Summary - A Call to Action

This paper outlines the requirements for spectrum archives necessary for data-driven policy decisions. There are nascent solutions that can be combined to assemble a usable data ecosystem, but efforts need to be taken in specific areas:

1) The community should focus on generating longitudinal, geographically and environmentally diverse data sets. A *minimum viable data product* to support policy decisions should be defined to inform data collection and storage.
2) A standardized metadata format for sensors and datasets should be adopted. We believe that extending SigMF and providing supporting infrastructure is the best solution.
3) RF data should be able to be stored and accessed in multiple data formats, including cloud-enabled storage. Rather than mandate a specific format, we believe that an "adaptation library" should be developed to access data and the specific data storage should be described though SigMF extensions.
4) We need *federated* data archives with unique data products identified through a DOI reference with metadata specific to RF data collection.
5) The community should develop query methods to search those federated datasets in a way that's usable and understandable by practitioners in the RF community. This should be coupled with data discovery methods.
6) Analysis and visualization tools should be developed to support policy making. Such tools should enable insights about past and present spectrum utilization, as well as supporting *what if* exercises. The tools should enable toggling between insights and raw data, and support exploration of uncertainty and trust in data products.

We believe this collections of standards will increase data reuse and begin to make that data suitable for data-driven policy. This in turn will support the proper instrumentation of field trials to enhance their policy impact by being able to reduce the uncertainty surrounding critically important considerations like aggregate interference which require the fusion of occupancy measurements with propagation modeling and receiver susceptibility.

REFERENCES

[1] M. Zheleva *et al.*, "Radio dynamic zones: Motivations, challenges, and opportunities to catalyze spectrum coexistence," *IEEE Communications Magazine*, 2023.

[2] M. K. Weldon, *The Future X Network: A Bell Labs Perspective*. CRC Press, 2016.

[3] Obama, Barack, "Presidential memorandum: Unleashing the wireless broadband revolution," https://obamawhitehouse.archives.gov/the-press-office/presidential-memorandum-unleashing-wireless-broadband-revolution, 2010, retrieved January, 2024.

[4] Federal Communications Commission, "600 MHz Band," https://www.fcc.gov/600-mhz-band, 2022, retrieved January, 2024.

[5] Federal Communications Commission, "Advanced Wireless Services," https://www.fcc.gov/wireless/bureau-divisions/broadband-division/advanced-wireless-services-aws, 2022, retrieved January, 2024.

[6] Federal Communications Commission, "3.5 GHz Band Overview," https://www.fcc.gov/wireless/bureau-divisions/mobility-division/35-ghz-band/35-ghz-band-overview, 2023, retrieved January, 2024.

[7] "White Space," https://www.fcc.gov/general/white-space, retrieved January, 2024.

[8] A. Chakraborty and S. R. Das, "Measurement-augmented spectrum databases for white space spectrum," in *CoNEXT*, 2014, p. 67–74.

[9] Federal Communications Commission, "Fact sheet - modernizing the 5.9 ghz band," https://docs.fcc.gov/public/attachments/DOC-367827A1.pdf, retrieved January, 2024.

[10] National Academies of Sciences, Engineering, and Medicine, *Analysis of Potential Interference Issues Related to FCC Order 20-48*. Washington, DC: The National Academies Press, 2023.

[11] M. G. Biggs *et al.*, "Measurements to Characterize Aggregate Signal Emmissions in the 2400-2500 MHz Frequency Range," https://its.ntia.gov/umbraco/surface/download/publication?reportNumber=95-323.pdf, 1995.

[12] E. D. Nelson, "LTE Uplink Aggregate Interference Measurement System," https://its.ntia.gov/umbraco/surface/download/publication?reportNumber=TM-21-552r.pdf, 2021.

[13] D. M. Le Vine *et al.*, "Aquarius RFI Detection and Mitigation Algorithm: Assessment and Examples," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 8, pp. 4574–4584, 2014.

[14] D. Boulware *et al.*, "An Analysis of Agregate CBRS SAS Data from April 2021 to January 2023," https://its.ntia.gov/umbraco/surface/download/publication?reportNumber=TR-23-567.pdf, 2023.

[15] National Institute of Standards and Technology, "WSRD Workshop: Making Data Available for National Spectrum Management," https://www.nist.gov/news-events/events/2023/05/wsrd-workshop-making-data-available-national-spectrum-management, 2023.

[16] F. H. Sanders *et al.*, "Broadband spectrum survey at San Diego, California," NTIA Technical Report, Tech. Rep. TR-97-334, Dec 1996.

[17] T. Taher *et al.*, "Global spectrum observatory network setup and initial findings," in *CROWNCOM*, 2014.

[18] Shared Spectrum Company, "General Survey of Radio Frequency Bands (30 MHz to 3 GHz)," http://www.sharedspectrum.com/wp-content/uploads/2010_0923-General-Band-Survey-30MHz-to-3GHz.pdf, Tech. Rep., 2010.

[19] A. Chakraborty *et al.*, "SpecSense: Crowdsensing for efficient querying of spectrum occupancy," in *INFOCOM*, 2017, pp. 1–9.

[20] D. Pfammatter *et al.*, "A Software-Defined Sensor Architecture for Large-Scale Wideband Spectrum Monitoring," in *IPSN*, 2015.

[21] A. Nika *et al.*, "Towards commoditized real-time spectrum monitoring," in *HotWireless*, 2014.

[22] S. Rajendran *et al.*, "Electrosense: Open and big spectrum data," *IEEE Communications Magazine*, vol. 56, no. 1, pp. 210–217, 2017.

[23] A. Scalingi *et al.*, "A framework for wireless technology classification using crowdsensing platforms," in *INFOCOM*, 2023.

[24] P. Vaccani, "Combining automated monitoring with a national licensing database for radio spectrum enforcement," in *National Symposium on Electromagnetic Compatibility*, 1989.

[25] C. Weber *et al.*, "Combining automatic radio monitoring and license databases for future spectrum management," in *IDAACS-SWS*, 2014.

[26] A. Nika *et al.*, "Empirical Validation of Commodity Spectrum Monitoring," in *SenSys*, 2016.

[27] CarrierIQ, "Understanding Carrier IQ Technology," Whitepaper, Dec 2011, available at http://www.franken.senate.gov/files/letter/111212_CarrierIQ_Attachment.pdf.

[28] "Shared commercial operations in the 3550-3650 MHz band," 80 F.R. § 36222, 2015.

[29] Josh Sanz, Ali Abedi, Anant Sahai, "Automatic indoor-outdoor detection using signals of opportunity," in *DySPAN*, 2024.

[30] K. Doke *et al.*, "VIA: Establishing the link between spectrum sensor capabilities and data analytics performance," in *INFOCOM*, 2024.

[31] T. Cooklev *et al.*, "The VITA 49 Analog RF-Digital Interface," *IEEE Circuits and Systems Magazine*, Fourthquarter 2012.

[32] "Ettus Research. Compressed header (CHDR) radio transport protocol," https://files.ettus.com/app_notes/RFNoC_Specification.pdf.

[33] "Signal Metadata Format (SigMF)," https://github.com/gnuradio/SigMF.

[34] "IEEE 802.22.3 - Standard for Spectrum Characterization and Occupancy Sensing," https://standards.ieee.org/develop/project/802.22.3.html.

[35] National Telecommunications and Information Administration (NTIA), "sigmf-ns-ntia," https://github.com/NTIA/sigmf-ns-ntia.

[36] The DOI Foundation, "The identifier DOI handbook," https://doi.org/10.1000/182.

[37] B. Rabenstein and J. Volz, "Prometheus: A next-generation monitoring system (talk)." Dublin: USENIX Association, May 2015.

[38] IQEngine, "IQEngine Toolkit," https://www.iqengine.org/, retrieved January, 2024.

[39] Microsoft Spectrum Observatory, "Spectrum Observatory," https://observatory.microsoftspectrum.com/, retrieved January, 2024.

[40] C. Brown and A. Ghasemi, "Evolution towards data-driven spectrum sharing: Opportunities and challenges," *IEEE Access*, 2023.

[41] C. Xu *et al.*, "Cosmic-conn: A cosmic-ray detection deep-learning framework, data set, and toolkit," *The Astrophysical Journal*, 2023.

[42] A. Chakraborty *et al.*, "Spectrum patrolling with crowdsourced spectrum sensors," in *INFOCOM*, Honolulu, HI, USA, 2018.