

Spectrum Monitoring/Panel 3: 5G Monitoring and Data Collection - The Feedback Loop

Presenter: Kaushik Chowdhury

Professor

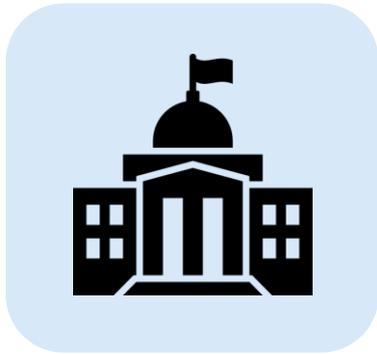
Associate Director, Wireless Institute of Internet of Things
Northeastern University, Boston MA, USA

krc@ece.neu.edu

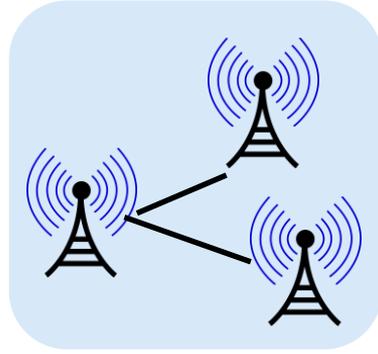
<http://genesys-lab.org>

Sources of 5G Datasets

FOCUS AREAS



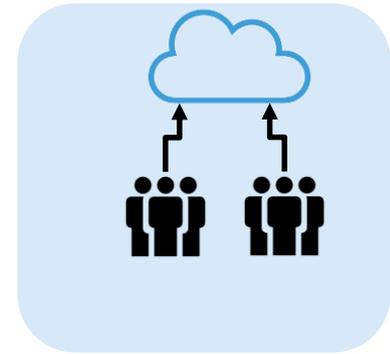
Institutional
(NTIA, NIST)



Experimental
(PAWR Platforms)



Emulation
(Colosseum)

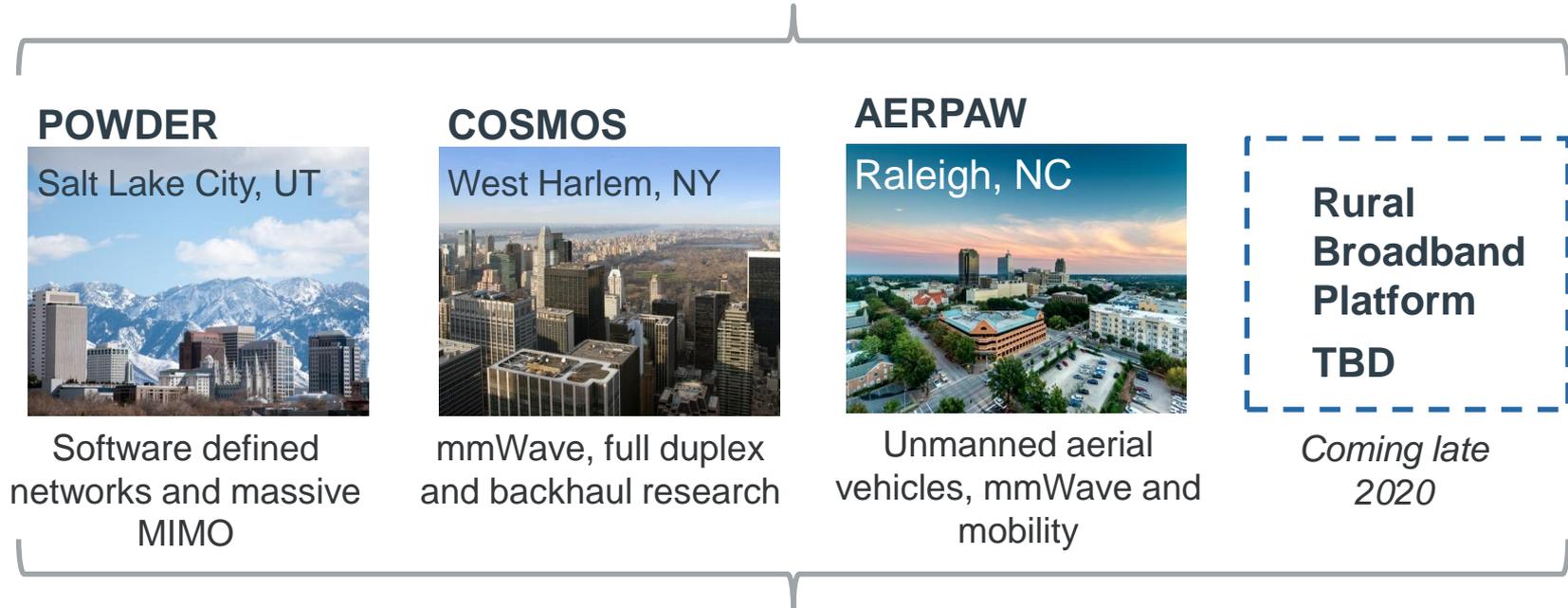


Crowdsourced
(Flightaware)

Takeaway: Need of committed integrator and dataset curator, shepherd

Focus #1: Experimental Dataset

PAWR PLATFORMS



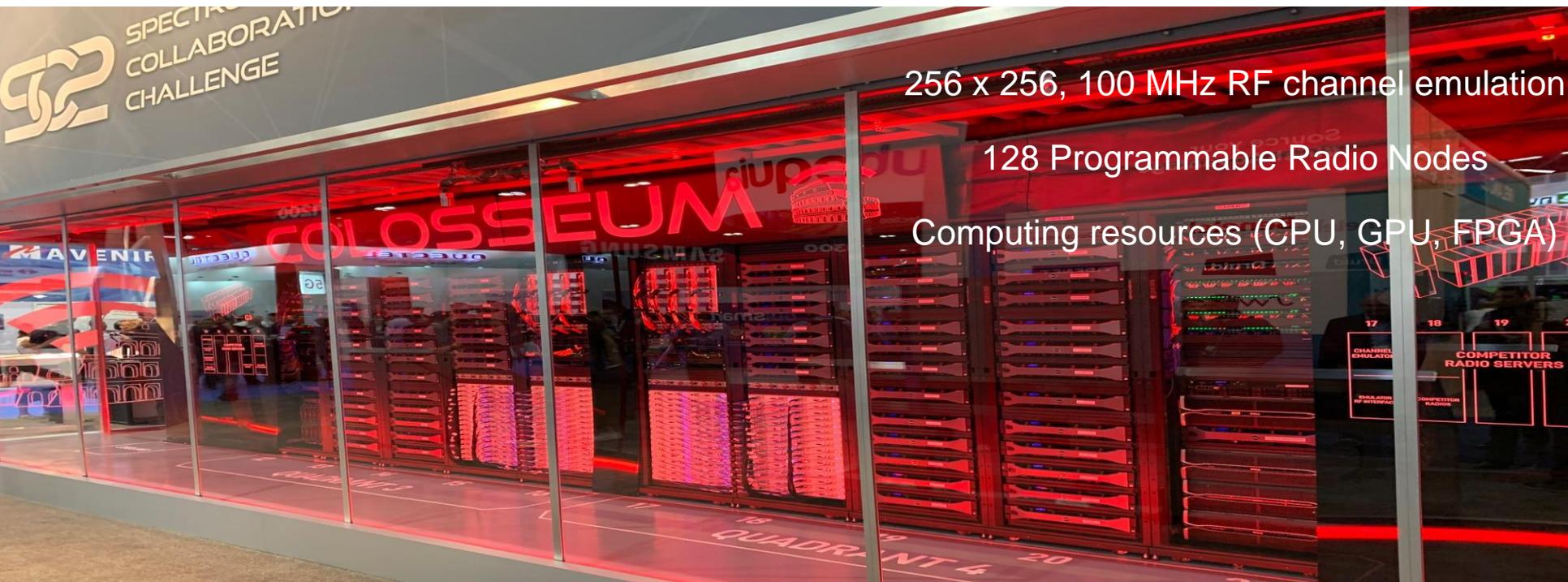
5G DATASETS

Takeaway: Need for reducing time from experiment conception to dataset creation

Focus #2: Emulated Dataset

COLOSSEUM : <https://www.northeastern.edu/colosseum/>

World's largest wireless network emulator with granularity at the RF signal level



256 x 256, 100 MHz RF channel emulation

128 Programmable Radio Nodes

Computing resources (CPU, GPU, FPGA)

Takeaway: Need for realistic scenarios, focus on relevant data

Sharing RF Datasets

Metadata representation: SigMF/SigMF-ntia

RF Fingerprinting Dataset Example:

<http://www.genesys-lab.org/oracle>

Dataset Description:

We are releasing two datasets **a) Dataset #1 : recordings of raw IQ samples collected from over-the-air transmissions of 16 USRP X310 transmitter radios ; b) Dataset #2: recordings of demodulated IQ symbols collected after equalizing over-the-cable transmissions of 16 IQ imbalance configurations.** In both the datasets, each recording consists of two files: a metadata file and a dataset file. The dataset file is a binary file of digital samples, and the metadata file contains information that describes the dataset. Our metadata and data format is an extension of, and compatible with [the SigMF specifications](#) .

- **Dataset #1** : It consists of recordings of collected raw IQ samples from 16, high-end X310 USRP SDRs with the same B210 radio as a receiver. The recordings are categorized into different folders with folder name "xxft", where xx represents the transmitter-receiver separation distance in feet. Each recording has a dataset file with an extension of *'sigmf-data'* , and a metadata file with an extension of *'sigmf-meta'*. These files are named in a specific format for more intuitive understanding. For example, the dataset file "**WiFi_air_X310_3123D7B_2ft_run1**" represents
 - **WiFi** :--> IEEE802.11a standard-compliant WLAN frame
 - **air** :--> medium of transmission
 - **X310** :--> the type of USRP radio
 - **3123D7B** :--> device serial ID
 - **2ft** :--> the transmitter-receiver separation distance in feet
 - **run1** :--> the recording number
 - **sigmf-data/sigmf-meta** :--> the extension of dataset file/metadata file

K. Sankhe, M. Belgiovine, F. Zhou, S. Riyaz, S. Ioannidis, and K. R. Chowdhury, "ORACLE: Optimized Radio Classification through Convolutional neural

Takeaway: Need for standardized data representation format



Thank You