RFDataFactory: One Stop Resource for RF-centric Datasets for the Wireless Community

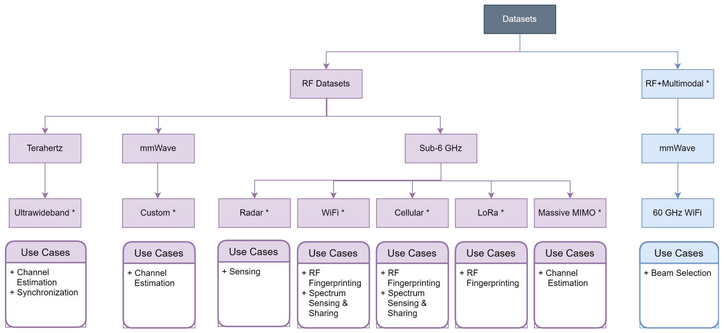
Award number:
CNS 2120447

www.rfdatafactory.com

Institute for the Wireless Internet of Things at Northeastern University

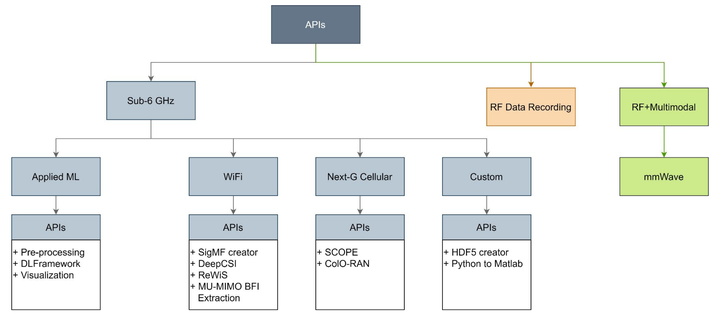
RICE

Wide Range of Tested Datasets



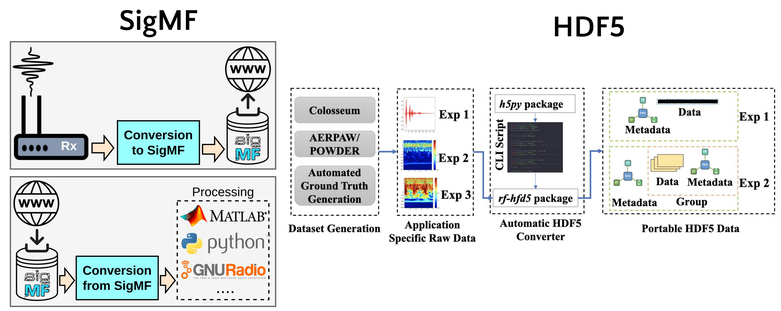
- A Total of **34** datasets are available at RFDataFactory website.
- **Dataset categories** include**:** Terahertz band (0.1-1THz), Millimeter-wave band (24-60GHz), sub-6GHz, and multimodal.

Software APIs to Create Datasets



- A Total of **15** APIs are available at RFDataFactory website.
- **API categories** include**:** applied ML, Wifi, Next-G Cellular, multimodal, and RF data recording.

Advocating RF Metadata Standards



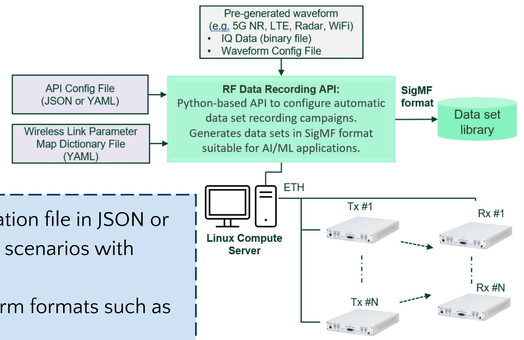
SigMF: An standard method for storing RF datasets, where each signal file (.bin) is accompanied with a metadata file (.json).

HDF5: Allows heterogenous data to be stored in a compact manner. File organization is stored as groups that may be embedded within data files, along with self-descriptive metadata for each file.

Spotlight #1. Data Collection API

Data Collection API:

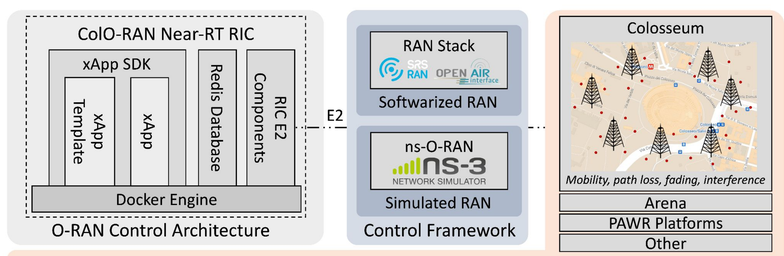
- Automates transmission and recording by interfacing NI's USRP devices via UHDs.



- Requires a single configuration file in JSON or YAML format, to configure scenarios with multiple Tx/Rxs
- Supports different waveform formats such as TDMS and MAT
- Stores datasets in the open source Signal Metadata Format (SigMF)

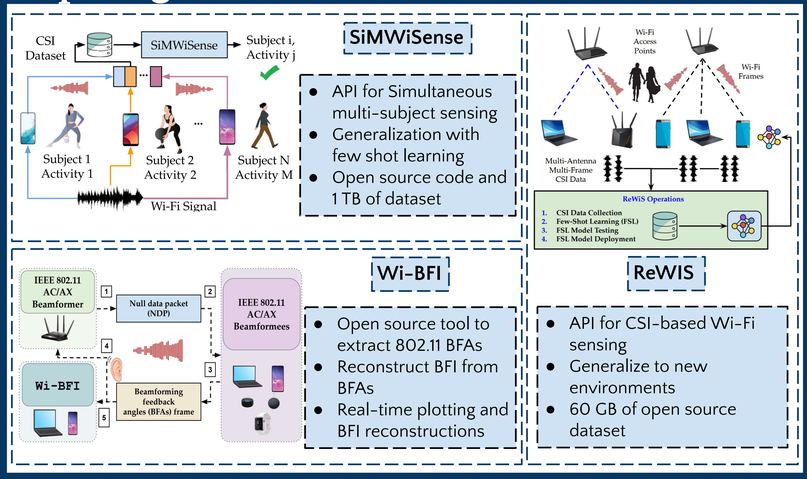
Designed in Collaboration with National Instruments

Spotlight #2. Data + RAN Control



- Open toolbox for data collection and experimentation with AI in O-RAN
- O-RAN-compliant near-real-time RIC, and different RAN softwares
- APIs for automatic data-collection and AI/ML RAN control

Spotlight #3. Wi-Fi based MU-MIMO



- API for Simultaneous multi-subject sensing
- Generalization with few shot learning
- Open source code and 1 TB of dataset

- Open source tool to extract 802.11 BFAs
- Reconstruct BFI from BFAs
- Real-time plotting and BFI reconstructions

- API for CSI-based Wi-Fi sensing
- Generalize to new environments
- 60 GB of open source dataset

Spotlight #4. Support for NSF PAWR

Public Deployment (mMIMO) Open Source Software for Dataset Collection

Agora

- mMIMO baseband computational performance
- Real-time PHY and MAC prototyping
  - Multi-User Scheduling (MAC)
  - Coding Schemes
  - Channel Estimation
  - Beamforming

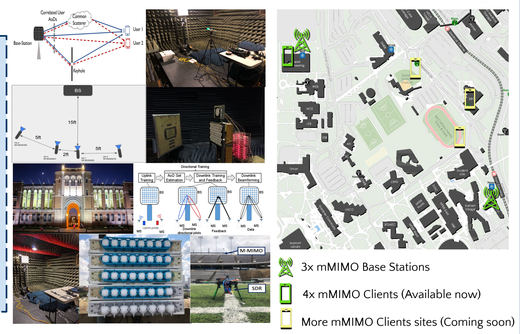
RENEWLab

- mMIMO uplink/downlink channel estimation
- mMIMO uplink/downlink data sample collection
- Waveform prototyping and evaluation
- SISO and uplink mMIMO PHY layer prototyping

9 mMIMO datasets (3600 website* users):

- Self-interference reduction technique for massive MIMO systems
- Comprehensive many-antenna MU-MIMO channel measurement campaign (indoor/outdoor)
- FDD Massive MIMO
- Multi-User MIMO dataset with inter-user channel correlation (AoD)
- Full-Duplex (Self-Interference in mMIMO)
- AoA (mMIMO)
- Uplink Massive MIMO with Coherent and Non-coherent Array
- Experimental Evaluation of AoA Estimation for UAV to Massive MIMO
- LensFD (full-duplex mMIMO)

*https://renew-wireless.org/datasets.html



3x mMIMO Base Stations

4x mMIMO Clients (Available now)

More mMIMO Clients sites (Coming soon)

Outreach: NSF Workshop

- **Participation summary:** 29 Academic, 13 Government/National Laboratory, and 8 Industry participants
- Live data collection demo
- Data collection on PAWR platforms

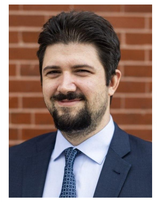
Discussion topics:

- Technical architectures for privacy protection in RF datasets
- RF dataset generation tools and resources
- Models for making RF datasets available to stakeholders
- International priorities and partnerships for large-scale RF data collection efforts
- Building public trust and confidence is using and certifying RF datasets

PI Team

Kaushik Chowdhury
Northeastern University
krc@ece.neu.edu

Tommaso Melodia
Northeastern University
melodia@northeastern.edu



Francesco Restuccia
Northeastern University
f.restuccia@northeastern.edu

Ashutosh Sabharwal
Rice University
ashu@rice.edu