



Data for
Refugees



D4R –The Data for Refugees Challenge: Sharing mobile CDR data for improving the conditions of Syrian refugees in Turkey

Albert Ali Salah^{1,2}, PhD

¹ Boğaziçi University, Department of Computer Engineering

² Utrecht University, Department of Information and Computing Sciences

February 17, 2019



Data for
Refugees

In a nutshell

- “Data for Refugees Turkey” is a Big Data innovation challenge, whereby Türk Telekom opens a large dataset of anonymized mobile phone usage to researcher groups for the purposes of providing better living conditions to Syrian refugees in Turkey.
- The Challenge is based on the successful Data for Development (D4D) series of challenges.
- Five areas are prioritized:
 - Safety & Security
 - Health
 - Education
 - Unemployment
 - Social Integration



Data for
Refugees

Who is involved?

- Supported by TUBITAK (Turkish Research and Scientific Council), universities, NGOs, and ministries related to the focus themes (i.e. health, education, and refugee matters). TUBITAK contact is Ali Görçin from TUBITAK BILGEM.
- The scientific team:
 - Albert Ali Salah, Boğaziçi Univ. & Utrecht Univ.
 - Alex ‘Sandy’ Pentland, MIT & Datapop Alliance
 - Emmanuel Letouzé, MIT & Datapop Alliance
 - Bruno Lepri, FBK
 - Patrick Vinck, Harvard Humanitarian Initiative & Datapop Alliance
 - Xiaowen Dong, Oxford Univ.
 - Yves-Alexandre de Montjoye, Imperial College London



Data for
Refugees

What was D4D?

- Data for Development Challenge, organized by France Telecom & Orange Cote d'Ivoire and the Government of Ivory Coast.
- 90 research groups participated.
- 267 applications, 166 of them received the **dataset**.
 - Collected from 5 million users, for the duration of 1 months.
 - Participants have to destroy their copy after the challenge.
- The best projects got an award + expenses paid to join NETMOB conference at MIT.



Data for
Refugees

Data for Refugees Dataset

- Collected from 200K refugees and 800K non-refugee residents, over the entire country, over one year.
- Includes:
 - Cell tower locations, lists of cell towers for each prefecture
 - Site-to-site antenna traffic on an hourly basis. Total number and duration of calls given, separated into “originating from refugees” and “not originating from refugees”.
 - Fine grained mobility of a small subset of (anonymous) users. Only 15 days for each user, at cell tower level.
 - Coarse grained mobility of a small subset of (anonymous) users. For the entire data collection period, but provided at prefecture level.



Data for
Refugees

Ethics

- A special ethics commission is formed, and evaluates all project proposals + all project reports.
- It provides feedback to proposes, and decides on one of four categories for each accepted project:
 - Candidate for Mention,
 - Publish without restriction,
 - Ask to consider adjustments,
 - Do not publish/Keep for restricted audience.



Data for
Refugees

D4R Project Evaluation Committee

Members from the Main Organizers

- Senem Özyavuz, Türk Telekom (chair)
- Fırat Yaman Er, Türk Telekom
- Yıldırım Bahadırlar, TÜBİTAK BILGEM
- Bülent Sankur, Boğaziçi University

Members from Academia

- Iyad Rahwan, MIT
- Alex Rutherford, MIT
- Geoffrey Fox, Indiana University
- Phuong Pham, Harvard Humanitarian Initiative



Data for
Refugees

D4R Project Evaluation Committee

Members from NGOs

- Anahi Ayala Iacucci, Internews
- Joséphine Goube, Techfugees
- Jean-Marie Garelli, UNHCR
- Mazen AboulHosn, IOM Turkey
- Claire Melamed, Global Partnership for Sustainable Development Data
- Ahmad Garibeh, Istanbul & I
- Nona Zicherman, UNICEF
- Manuel Garcia-Herranz, UNICEF
- Vedran Sekara, UNICEF

Members from the Government

- Mithat Büyükhan, TR Ministry of Education, Lifelong Learning General Directorate
- Ömer Hakan Şimşek, TR Ministry of Health, IT
- Mustafa Yılmaz, TR Ministry of Interior, Directorate General of Migration Management



Data for
Refugees

D4R Main Purposes

- Contribute to the welfare of the refugee populations,
- Gain insights on key issues, including safety and security, health, education, unemployment, social integration and segregation, mobility, and distribution of resources and infrastructure,
- Help governments and international bodies model the dynamics of the refugee populations and to discover vulnerabilities (socio-economic vulnerabilities, gaps in education and services, etc.),
- Seed further projects, co-created with refugees, resulting in new applications, services, and innovative solutions for refugees in Turkey and elsewhere.



Data for
Refugees

Priority Questions in D4R

- Safety & Security
 - Violence, theft, illegal trafficking
- Health
 - Access to resources, spread of diseases, vaccination
- Education
 - Access to education, language learning, schools
- Unemployment
 - Movement due to unemployment, skill and resource management
- Social Integration
 - Events and institutions for social integration, segregated/mixing patterns of behavior



Data for
Refugees

Evolution of D4R

- January 2017: Conception of the idea (Boğaziçi & TÜBİTAK)
- April 2017: Formation of the Scientific Committee
- April 2017: Türk Telekom approached with the idea
- July 2017: Türk Telekom approves data description
- September 2017: Formation of the Project Evaluation Committee starts, UNHCR contacted for support
- October 2017: University ethics board approval



Data for
Refugees

Evolution of D4R

- October 2017: IOM and UNICEF contacted for support
- December 2017: Boğaziçi-MIT MISTI project accepted
- January 2018: Website goes public, UNHCR support
- March 2018: IOM, Ministry of Interior, UNICEF support
- April 2018: 104 proposals received and screened, 61 teams receive the dataset
- July 2018: 31 reports submitted
- 21 January 2019: Closing workshop in Istanbul
- March 2019: Planned publication of D4R white paper
- Mid 2019: Planned publication of D4R book



Data for
Refugees

Key issues for UNHCR

- The sharing of the data must be permitted by the owners of the data, i.e., the refugees
- The data should be anonymized in a way to make it impossible to identify individuals in it
- The data should not allow other governments to track specific refugees



Data for
Refugees

Project preconditions

- Participation will be under a strict agreement to protect data confidentiality and privacy.
- Research teams applying for the Challenge will be lead by a permanent staff member of an institution, and all team members will be individually named in the agreement.
- A short proposal describing the aims of the team will be submitted for application.
- The applications will be evaluated primarily **on their ultimate benefit to the refugees**. Commercial use of data is not permitted.



Data for
Refugees

Denied projects

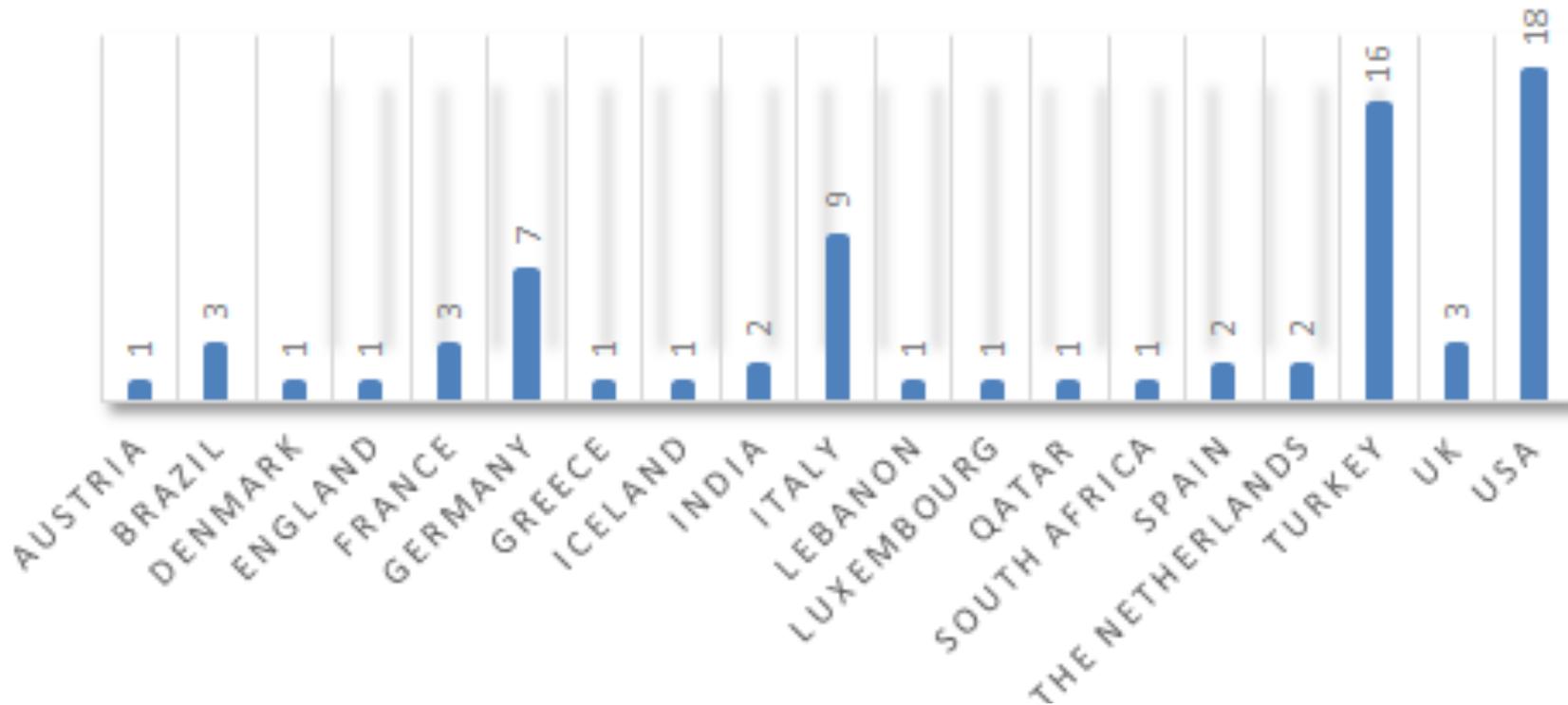
- Projects with obvious commercial interest
 - Apps targeted for refugees
 - Data science startups hungry for data
- Projects with no/poor technical description
 - Project proposing to find refugee ‘influencers’
 - Project proposed for raising awareness
- Projects with potentially sensitive topics
 - Defense related groups (government or otherwise)



Data for
Refugees

Projects

PARTICIPATING INSTITUTIONS ACCORDING TO COUNTRY





Data for
Refugees

Denied results

- Reports mentioning the ruling political party explicitly (e.g. migration effects on voting outcomes)
- Reports mentioning illegal work on certain areas (e.g. construction sites run by companies aligned to the government)
- Reports mentioning third parties explicitly (e.g. hospitals where refugees are getting services)



Data for
Refugees

Results

- All reports available for downloading at:
 - <http://d4r.turktelekom.com.tr/Content/Documents/d4r-proceedings.pdf>
- White paper in Turkish being prepared for government, policy makers and other stakeholders
- Edited book '[Guide to analytics in refugee scenarios](#)', to appear from Springer Int.



Data for
Refugees

ETHICS CLAUSES OF D4R



Data for
Refugees

Retention, destruction and archiving

The Challenge mandates that all participants destroy the Database upon completing the challenge. Any publication based on the Database requires the pre-approval of the Project Evaluation Committee. Any further use of the Database (for instance to complete numerical experiments for a publication under review) will be regulated by PEC, and extensions are to be conditionally granted for specific purposes on a case by case basis.



Data for
Refugees

Information

Every care has been given to ensure that the information provided in the Database does not cause any harm, prejudice, or distress to customers, regardless of their refugee status. PEC provides an additional layer of control, and will examine project reports confirm to this maxim.



Data for
Refugees

Access, correction, erasure, objection

The Database does not contain personal information, and it is not possible for individuals to request access to personal data. The data are anonymised and aggregated in a way to prevent identification of persons. For the same reason, correction, erasure, and objection do not apply for the Database. We note that this is a stricter protection condition than most envisioned CDR usage scenarios. Furthermore, the Database does not contain children's data, as each registered customer has to be over 18 years of age, and thus legally permitted to own and use a mobile phone line. Personal profiling (such as used for CRM applications) is not possible with the Database.



Data for
Refugees

Responsibility and accountability

The responsibilities of all parties concerned are defined clearly, and set out in the data agreement prepared by TT lawyers.



Data for
Refugees

Data protection by design and default

Data collection follows this principle, where any names, real phone numbers, or other identifying information is excluded from the design of the Database. The pseudo-random numbers representing customers are not stored anywhere along with actual phone numbers. Subsequently, the anonymisation works only one way. Refugee status is indicated by purposefully noisy indicators, and no effort is spent to ensure its validity. Subsequently, only aggregate-level conclusions can be drawn from the Database. It is not possible to use the Database for the surveillance and tracking of individuals.



Data for
Refugees

Limitations

Data access is not provided to institutions in a blanket permission, but to specific individuals within institutions, whose names and roles in the proposed research project is clearly defined in the data agreement.



Data for
Refugees

Collaboration

- We have:
 - Computer analysis of multimodal multilevel human behavior
 - Operational experience on a data challenge on refugees
- We look for:
 - Partners with interesting projects we can help realize, with social impact
 - Especially work with children and refugees
- Contact:
 - Albert Ali Salah, a.a.salah@uu.nl
 - Twitter: @SzassTam