GDPR Analysis of the BigBang Project

# Clients: Niels ten Oever, Christoph Becker

Dear Sir or Madam,

We are writing to you on behalf of the academic quantitative data analysis project BigBang.

We study the governance of transnational communication networks by analyzing email and standards archives. That infrastructure increasingly underpins modern information societies and is governed through standards-setting in private governance fora. Our software, which is licensed under an open source license (MIT), allows for the downloading of emails from the public email archives of standard-setting bodies and subsequently allows for doing network, discourse, and statistical analysis.

In future versions of this software we would like to make the email data directly available to other researchers in order to reduce the strain on email servers of standards-setting bodies and to enhance accessibility and usability for researchers.

We would like to solicit pro bono advice to ensure that we or our users are not violating the EU’s General Data Protection Regulation (GDPR).

Questions:

1. **Are the collecting and processing of personal data via BigBang compliant with the EU’s General Data Protection Regulation (GDPR)?**

BigBang allows for the scraping of email archives of various Internet standard-setting organizations, including 3GPP, W3C, IETF, ICANN, and RIPE. Emails are being downloaded either as local copies to local machines or to a shared server in Amsterdam that is controlled and administered (with password protections and appropriate data security measures) by one of the clients.

Examples of such emails are provided in Appendix A.

BigBang allows for various forms of data processing to aggregate and analyze emails. BigBang’s current capabilities are mainly focused on header analysis. As you can see from the examples in the Appendix, the relevant information includes inter alia time and timezones, sender email address, sender user names, recipient email addresses, recipient user names, and user affiliation (as revealed by the email domain name or email signature). BigBang also allows for text body analysis but is currently limited to word search (filtering emails by the use of a certain word). In future, we might add more sophisticated forms of text analysis.

The analysis enabled by BigBang can be rendered in graphs and can be used for the writing of academic articles. Civil society organizations may also use BigBang to interrogate standard-setting processes and to render them more accountable by writing reports, news articles, blog posts etc.

We understand that the emails contained in standard-setting bodies’ archives, despite being publicly available, are to be considered “personal data” under the EU’s General Data Protection Regulation (GDPR) and that the analysis facilitated by BigBang is to be considered “processing” of personal data.

We would like to argue that BigBang enabled processing of personal data is necessary for the purposes of legitimate interests pursued by researchers and civil society organizations under Article 6(1)(f) GDPR. We do not believe that the public interest in the transparency and accountability of private standard-setting organizations is overridden by the interests or fundamental rights and freedoms of the data subject implicated by the email archives that BigBang mines.

Pseudonomizing or anonymizing the datasets is not only technically challenging but would also adversely affect the dataset’s value as a research and accountability tool.

1. **How can BigBang make its datasets available to other researchers without violating GDPR?**

We are contemplating to make aggregated datasets available to other researchers in order to reduce the strain on email servers of standard-setting organizations and to improve the accessibility and usability of BigBang for researchers (scraping the emails from the archives takes time and requires a certain amount of technical expertise as well as storage capacity).

We would like to offer two ways of accessing the aggregated datasets: a) a possibility to download the dataset; b) a possibility to process the data on our server.

We understand that either possibility would be considered “processing” of personal data under the GDPR. How can we offer these possibilities without violating the GDPR?

We would like to require researchers to file a data access permit application. A template of such an application is available at <https://github.com/datactive/bigbang/blob/main/data_access_permit_application.md> and provided in Appendix B. Are these terms sufficient or would you recommend additional or different terms to facilitate researcher-to-researcher sharing of the datasets?

As stated above, pseudonomizing or anonymizing the datasets is not only technically challenging but would also adversely affect the dataset’s value as a research and accountability tool.

Thank you very much for considering this request. We look forward to hearing from you and would be happy to answer any questions you may have.

Sincerely,

Signatures

# 

# 

# Appendix A: Data Sample

### Message Nr1:

From MAILER-DAEMON Thu Aug 26 16:49:36 2021

Content-Transfer-Encoding: quoted-printable

MIME-Version: 1.0

message-id: CY4PR10MB15280221A838A614A5E5D7C0B3130@CY4PR10MB1528.namprd10.prod.outlook.com

sender: 3GPP\_TSG\_RAN\_WG4\_NR-MIMO-OTA 3GPP\_TSG\_RAN\_WG4\_NR-MIMO-OTA@LIST.ETSI.ORG

subject: 'Revisions to TPs on FR2 channel verification and FR2 DoT'

from: "Rodriguez-Herrera, Alfonso" Alfonso.Rodriguez-Herrera@SPIRENT.COM

date: Thu, 20 Feb 2020 17:40:30 +0000

Content-Type: text/plain; charset="utf-8"; Content-Type="multipart/mixed"; MIME-Version="1.0"

reply-to: "Rodriguez-Herrera, Alfonso" Alfonso.Rodriguez-Herrera@SPIRENT.COM

Archived-At: <https://list.etsi.org/scripts/wa.exe?A2=3GPP\_TSG\_RAN\_WG4\_NR-MIMO-OTA;5be0a403.2002C&S=>

Dear NR MIMO OTA Colleagues,

Please find attached an advanced copy of the TPs for FR2 DoT, and FR2 channel=

verification. The previously submitted TPs did not have the "Track Changes" =

turned on, so it is difficult to see what changes are being proposed.

The attached are expected to be submitted as revised TPs.

Apologies for the inconvenience.

Regards,

Alfonso

### Message Nr2:

From MAILER-DAEMON Thu Aug 26 16:49:38 2021

Content-Transfer-Encoding: quoted-printable

MIME-Version: 1.0

sender: 3GPP\_TSG\_RAN\_WG4\_NR-MIMO-OTA 3GPP\_TSG\_RAN\_WG4\_NR-MIMO-OTA@LIST.ETSI.ORG

subject: 'Re: draft WF on FR2 MIMO OTA'

from: Ruixin Wang wangruixin@CAICT.AC.CN

date: Fri, 22 Nov 2019 12:14:47 -0800

message-id: 2019112212144537800715@caict.ac.cn

Content-Type: text/plain; charset="utf-8"; MIME-Version="1.0"; Content-Type="multipart/mixed"

reply-to: Ruixin Wang wangruixin@CAICT.AC.CN

Archived-At: <https://list.etsi.org/scripts/wa.exe?A2=3GPP\_TSG\_RAN\_WG4\_NR-MIMO-OTA;6040f53a.1911D&S=>

Hi All,

offline comments received. Please find attached version. Your comments are we=

lcome.

Please let me know if you want to co-sign.

Ruixin

CAICT

=20

From: Ruixin Wang

Date: 2019-11-22 10:39

To: 3GPP\_TSG\_RAN\_WG4\_NR-MIMO-OTA

Subject: draft WF on FR2 MIMO OTA

Hi All,

Please find attached draft WF on FR2 MIMO OTA.

Ruixin

CAICT

### Message Nr3:

From MAILER-DAEMON Thu Aug 26 16:49:40 2021

Content-Transfer-Encoding: quoted-printable

MIME-Version: 1.0

sender: 3GPP\_TSG\_RAN\_WG4\_NR-MIMO-OTA 3GPP\_TSG\_RAN\_WG4\_NR-MIMO-OTA@LIST.ETSI.ORG

date: Fri, 22 Nov 2019 19:12:11 +0000

reply-to: Ting-Wei Kang =?utf-8?b?5bq35bqt57at?=ting-wei.kang@MEDIATEK.COM

message-id: 51d91738e2814d1c85bb6233718f0f5f@mtkmbs02n1.mediatek.inc

subject: 'Re: draft WF on FR2 MIMO OTA'

from: Ting-Wei Kang =?utf-8?b?5bq35bqt57at?=ting-wei.kang@MEDIATEK.COM

in-reply-to: 2019112210395147521610@caict.ac.cn

Content-Type: text/plain; charset="utf-8"; Content-Type="multipart/mixed"; MIME-Version="1.0"

comments-to: Ruixin Wang wangruixin@CAICT.AC.CN

Archived-At: <https://list.etsi.org/scripts/wa.exe?A2=3GPP\_TSG\_RAN\_WG4\_NR-MIMO-OTA;7996391e.1911D&S=>

Hi Ruixin,

We add =E2=80=9CRevisit Initial phase definition of channel model, that is co=

mmon part for FR2 and FR1.=E2=80=9D in page 2. Please consider it.

Best,

TingWei

From: 3GPP\_TSG\_RAN\_WG4\_NR-MIMO-OTA [mailto:3GPP\_TSG\_RAN\_WG4\_NR-MIMO-OTA@LIST.=

ETSI.ORG] On Behalf Of Ruixin Wang

Sent: Saturday, November 23, 2019 2:40 AM

To: 3GPP\_TSG\_RAN\_WG4\_NR-MIMO-OTA@LIST.ETSI.ORG

Subject: draft WF on FR2 MIMO OTA

Hi All,

Please find attached draft WF on FR2 MIMO OTA.

Ruixin

CAICT

# Appendix B

# **BigBang - Data Access Permit Application**

## **Commitments**

Every person named below makes this statement of commitment.

* I shall not hand over the data to outside parties.
* I shall store the data in a safe and secure manner.
* If I become aware of a data security violation associated with the data I shall report this to the BigBang mailinglist.
* I will treat the data in accordance and in compliance with the licenses of the original data providers, such as the 3GPP, IETF, ICANN, RIPE, and others.
* This data should be treated in accordance with the Ethical Principles and Guidelines for the Protection of Human Subjects of Research as outlined in the [Belmont Report](https://www.hhs.gov/ohrp/sites/default/files/the-belmont-report-508c_FINAL.pdf) (Department of Health, Education, and Welfare; 1978).

## **Applicant**

Surname:

First name:

Email:

Research institute, educational establishment or other facility:

Title:

Degree:

## **Other Research Group Members**

Surname:

First name:

Email:

Research institute, educational establishment or other facility:

Title:

Degree:

## **Research/Application Description**

Name & Subject of the research/application (short description):

Has BigBang granted an earlier permit:

[ ] Yes, by whom (name) and when (date):

[ ] No

# Appendix C: Relevant Links

GitHub repository: <https://github.com/datactive/bigbang>

BigBang documentation: <https://bigbang-py.readthedocs.io/en/latest/>

BigBang website: <http://datactive.github.io/bigbang/>

Client website: <https://nielstenoever.net>

# 