**DGB Public Use File Briefing Document**

**PUF Name:** FHIR Synthetic Public Use File (FHIR-SynPUF).

**Office/Center:** Office of Enterprise Data and Analytics (OEDA)

**Executive Sponsor:** Chris Cox, Director, OEDA

**Component Point of Contact:** Ronnie Tan

1. **Briefly describe the purpose of the PUF**.

The purpose of the FHIR-SynPUF is to provide de-identified synthetic test data to assist software developers in the development, testing, and demonstration of applications that would connect to the CMS Blue Button Application Programming Interface (API). The Blue Button API will provide a developer-friendly, standards-based data API that enables beneficiaries to connect their Medicare claims data to the applications, services, and research programs they trust. CMS is proposing to make de-identified FHIR synthetic test data publicaly available so that software developers interested in building applications that will connect to the CMS Blue Button API can understand the data formats and specifications that their applications will need to utilize.

1. **Briefly describe the contents of the PUF**.

The FHIR-SynPUF is a synthetic data file constructed from Medicare beneficiary and claims data for a cohort of 30,000 beneficiaries selected from the years 1999, 2000, and 2014. It contains Medicare Part A inpatient claims, Part B carrier claims, Part D prescription event records, and a Beneficiary Summary File. The FHIR-SynPUF is available as 4 separate data files (one per file type).

1. **Check the applicable category below for the PUF:**

This PUF is for a single data product.

This PUF is for a series of releases with substantially the same content – specify the interval at which future products will be released:

This PUF is for the re‑release of an approved product, with the addition of previously unreleased data – describe the new data:

1. **Briefly describe the steps you’ve take to validate the accuracy of the data in the PUF.**

The FHIR-SynPUF is not a study or research file. The data were significantly altered and are not intended to be an accurate representation of Medicare beneficiaries and their claims. Data distributions observed in the FHIR-SynPUF will differ from true distributions in the source data and correlations will not be maintained. The FHIR-SynPUF primarily is useful for learning how to work with the types of Medicare data that will be available via the Blue Button API, but not to gather any reliable quantitative understanding of the clinical, demographics, or epidemiology of the Medicare population. These limitations will be clearly conveyed to users of the FHIR-SynPUF.

1. **The Privacy Act governs the collection, maintenance, use, and dissemination of individually identifiable information maintained in systems of records by federal agencies**.

**1. Does the PUF include individually identifiable information?**

☐ Yes – please answer question 2

X No – skip to Section G

**2. Is there a Privacy Act System of Records (SOR) from which the information used to create the PUF will be retrieved[[1]](#footnote-1), including a routine use that permits the creation of a PUF?**

☐ Yes – please list:

X No

1. **The HIPAA Privacy Rule permits the disclosure of protected health information that has been de-identified. How does the PUF meet the de-identification standard in §164.514 of the HIPAA Privacy Rule?**

☐ Safe Harbor (the file does not contain any of the 18 identifiers listed at §164.514(b)(2)[[2]](#footnote-2) AND you are not aware of any other publicly available data that could be linked with the data in the PUF and used to re-identify individuals)?

X Expert Determination (per §164.514(b)(1)) please attach the expert’s evaluation of the risk of the file.

☐ PUF is not developed from protected health information

☐ Other – please explain:

1. **CMS privacy policy stipulates that no cell (e.g. admittances, discharges, patients, services) 10 or less may be displayed. Does your PUF contain any cells of 10 or less?**

The PUF does not contain small cells and includes supplementary suppression to ensure that users cannot recalculate suppressed information.

The PUF has been statistically certified as de-identified even through it contains small cells (please attach the statistical certification).

The PUF does not contain any information on individuals.

Other – please explain:

1. **Are you aware of any other publicly available data that could be linked with the data in the PUF and used to re-identify individuals?**

Yes – please list:

No

N/A – please explain:

1. **What is the PUF dissemination plan, including the proposed website location?**

The FHIR Syn-PUF will be posted on the public HHS github site: <https://github.com/HHS>.

1. **Who is the intended audience for the PUF? If the PUF has been released previously, how many times has it been viewed or downloaded?**

This de-identified test data are intended for use by third-party Blue Button software developers, to assist them in the development, testing, and demonstration of their applications. By leveraging this public resource, developers will be able to better provide applications that improve Medicare beneficiary’s access to their own health data.

1. If you need assistance regarding SOR information, please contact the Division of Security, Privacy Policy & Governance at [ciso@cms.hhs.gov](mailto:ciso@cms.hhs.gov) [↑](#footnote-ref-1)
2. Names, geographic subdivisions smaller than a state, elements of dates (except year) for dates that are directly related to an individual, telephone numbers, vehicle identifiers and serial numbers, fax numbers, device identifiers and serial numbers, email addresses, Web Universal Resource Locators (URLs), social security numbers, Internet Protocol (IP) addresses, medical record numbers, biometric identifiers, health plan beneficiary numbers, full-face photographs and any comparable images, account numbers, any other unique identifying number, characteristic, or code [↑](#footnote-ref-2)