

## **DATA USAGE AGREEMENT FOR THE BERLIN-TÜBINGEN-ONCOLOGY CORPUS (BRONCO150)**

BRONCO150 is a corpus containing selected and scrambled sentences of 150 German clinical discharge summaries of cancer patients treated at Charité – Universitätsmedizin Berlin or Universitätsklinikum Tübingen. All discharge summaries were manually anonymized. The process was approved by the respective data protection offices at both clinics. The corpus was manually annotated with the following types of entities and normalized to terminologies given in brackets: diagnosis (ICD10), treatment (OPS), medication (ATC), and whether those entities are related to a certain laterality, are negated, speculated or will merely be possible in the future.

The corpus was created by the nationally funded project “Personalizing Oncology via Semantic Integration of Data” (PersOnS), see <https://persons-project.informatik.uni-tuebingen.de/>.

BRONCO150 is provided by the Corpus Provider

Prof. Dr. Ulf Leser  
Wissensmanagement für Bioinformatik (WBI)  
Institut für Informatik  
Humboldt-Universität zu Berlin  
Unter den Linden 6, 10099 Berlin  
Tel. +49 (0)30 2093 3901  
Fax. +49 (0)30 2093 5484  
E-mail [leser@informatik.hu-berlin.de](mailto:leser@informatik.hu-berlin.de)

Prof. Dr. Ulrich Keilholz  
Direktor des Charité Comprehensive Cancer Center (CCCC)  
Charité – Universitätsmedizin Berlin  
Charitéplatz 1, 10117 Berlin

Prof. Dr. Nisar Peter Malek  
Ärztlicher Direktor Innere Medizin I  
Universitätsklinikum Tübingen  
Ottfried-Müller-Straße 10, 72076 Tübingen

represented by Prof. Dr. Ulf Leser

to the Data User (as signed below) under the following terms:

1. BRONCO150 may only be used for non-commercial research on information extraction from clinical documents within a certified research organization.
2. Usage of BRONCO150 is granted to individual Data Users only. All prospective Data Users of BRONCO150 must fill out this data usage agreement individually.
3. The Data User agrees that he/she will (i) not attempt to re-build original discharge summaries or parts thereof from BRONCO150 and (ii) not attempt to identify or re-identify individual persons, hospitals or doctors from BRONCO150.
4. Data User certifies that his or her copy of BRONCO150 is stored only in a single copy on computers under administration of the Data User. Data User certifies that he or she will take proper action for protecting this copy from being accessed, read or copied by any non-authorized person.
5. The Data User agrees to delete BRONCO150 after twelve months signing this agreement. The Corpus Provider must be informed that deletion of the corpus by the Data User has been done. An extension of data usage is possible by signing this agreement again.
6. BRONCO150 is provided free of charge.
7. BRONCO150 comes with absolutely no warranties including (but not limited to) the correctness of the information provided in the text corpus itself.
8. The Data User will not disclose, disseminate, or otherwise share BRONCO150 to or with any other person or entity, including any subcontractor, for any purpose. BRONCO150 must not be transmitted electronically to other services not under administration of the Data User, such as online translation services.
9. In no event shall the Corpus Provider be liable to the Data User for direct, indirect, special, incidental, punitive or consequential damages of any kind arising in any way out of this agreement, rights granted herein or by the use of BRONCO150.
10. BRONCO150 in all (written) forms, shall be and remain the property of the Corpus Provider, represented by Prof. Dr. Ulf Leser.
11. Data User will provide with this document the Corpus Provider with a short summary (less than 100 words) describing the purpose of his or her research based on BRONCO150. Data User agrees that all his or her actual research activities with BRONCO150 will adhere to this description. Using BRONCO150 for a different kind of research requires signing a new data usage agreement with a new description.
12. Data User agrees that his or her name, contact information, and the research summary are stored in electronic form by the Corpus Provider. This information will be used to (a) inform Data Users when updates of BRONCO150 are available and to (b) create proper anonymized corpus distribution statistics. Additionally, the information might be used to track violations of this agreement. This information will not be shared with any other person or party. It will be deleted once this Data Usage Agreement is expired or cancelled by the Data User or by the Corpus Provider.
13. Contributions which are based on BRONCO150 must cite the following publication:

*Madeleine Kittner, Mario Lamping, Damian T Rieke, Julian Götze, Bariya Bajwa, Ivan Jelas, Gina Rüter, Hanjo Hautow, Mario Sanger, Maryam Habibi, Marit Zettwitz, Till de Bortoli, Leonie Ostermann, Jurica Ševa, Johannes Starlinger, Oliver Kohlbacher, Nisar P Malek, Ulrich Keilholz, Ulf Leser (2021). "Annotation and initial evaluation of a large annotated German oncological corpus", JAMIA Open, Volume 4, Issue 2, ooab025.*

I hereby apply for a copy of BRONCO150 and agree to the above terms as Data User.

Name: \_\_\_\_\_

Affiliation: \_\_\_\_\_

Position: \_\_\_\_\_

Email: \_\_\_\_\_

I will use BRONCO150 for the following research purpose (< 100 words)

---

---

---

---

---

---

---

---

---

---

---

The following person (typically the group leader) is responsible for the secure administration and management of the computer system on which BRONCO150 will be stored

Name: \_\_\_\_\_

Affiliation: \_\_\_\_\_

Position: \_\_\_\_\_

Email: \_\_\_\_\_

Date: \_\_\_\_\_

Signature of Data User: \_\_\_\_\_

Signature of responsible person: \_\_\_\_\_