

# Supplementary Materials

## S1. PHI Tag Inventory

Table S1 lists the 24 PHI categories used for annotation and model training, grouped by semantic type.

Table S1: Complete inventory of the 24 PHI tag categories used for annotation and BIO-format model training, grouped by semantic type. English glosses are provided for reference.

| Group                       | Tags                                                                                                                                                                               |
|-----------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Person & demographics       | <PERSOON> (person)<br><PERSOONAFKORTING> (initials)<br><LEEFTIJD> (age)                                                                                                            |
| Institution & location      | <ZIEKENHUIS> (hospital)<br><ADRES> (address)<br><PLAATS> (place)<br><BEDRIJF> (company)                                                                                            |
| Study-related               | <STUDIENAAM> (study name)                                                                                                                                                          |
| Temporal                    | <DATUM> (date)<br><TIJD> (time)                                                                                                                                                    |
| Contact & web               | <TELEFOONNUMMER> (phone no.)<br><EMAIL> (email)<br><URL> (URL)                                                                                                                     |
| Institutional identifiers   | <PATIENTNUMMER> (patient no.)<br><ZNUMMER> (Z-number)<br><RAPPORT\ ID> (report ID)<br><DOCUMENTID> (document ID)<br><DOCUMENTNUMMER> (document no.)<br><PHINUMMER> (other PHI no.) |
| Dutch/regulated identifiers | <ACCREDITATIE\ NUMMER> (accreditation no.)<br><IBAN> (IBAN)<br><BSN> (citizen no.)<br><BIGNUMMER> (BIG no.)<br><AGBNUMMER> (AGB no.)                                               |

## S2. Per-tag Performance by Dataset

This section provides a full breakdown of merged-entity detection performance for each PHI tag, reported separately for each dataset subset. These results complement the aggregated per-tag table in the main text (Table 2) by showing the extent to which performance varies across report types and institutions. Tags that do not appear in a given dataset are omitted from that subset’s block.

Table S2: Per-tag performance by dataset for the merged-entity evaluation. Dutch tag names are shown with concise inline English glosses. Detection recall counts a gold entity as detected when every token in the annotated span receives a non-0 label, irrespective of the assigned entity type or whether the predicted span extends beyond the gold span, whereas precision, recall, and F1-score require the exact correct entity label.

| Dataset                             | Tag                                                 | N   | Det. | rec. | Prec. | Rec. | F1   |
|-------------------------------------|-----------------------------------------------------|-----|------|------|-------|------|------|
| <b>RUMC pathology</b> ( $N = 199$ ) |                                                     |     |      |      |       |      |      |
|                                     | accreditatie<br>nummer ( <i>accreditation no.</i> ) | 111 | 1.00 |      | 1.00  | 1.00 | 1.00 |
|                                     | datum ( <i>date</i> )                               | 571 | 1.00 |      | 0.99  | 0.99 | 0.99 |
|                                     | persoon ( <i>person</i> )                           | 99  | 0.98 |      | 0.86  | 0.90 | 0.88 |
|                                     | persoonafkorting ( <i>initials</i> )                | 427 | 0.98 |      | 0.94  | 0.95 | 0.95 |
|                                     | phinummer ( <i>other PHI no.</i> )                  | 4   | 1.00 |      | 0.00  | 0.00 | 0.00 |
|                                     | rapport id ( <i>report ID</i> )                     | 198 | 0.98 |      | 0.98  | 0.97 | 0.98 |
|                                     | tijd ( <i>time</i> )                                | 3   | 1.00 |      | 0.75  | 1.00 | 0.86 |
|                                     | url ( <i>URL</i> )                                  | 7   | 1.00 |      | 1.00  | 1.00 | 1.00 |
|                                     | ziekenhuis ( <i>hospital</i> )                      | 107 | 0.99 |      | 0.86  | 0.91 | 0.88 |
| <b>RUMC radiology</b> ( $N = 542$ ) |                                                     |     |      |      |       |      |      |
|                                     | adres ( <i>address</i> )                            | 4   | 0.75 |      | 0.33  | 0.25 | 0.29 |
|                                     | datum ( <i>date</i> )                               | 571 | 1.00 |      | 0.98  | 0.95 | 0.96 |
|                                     | leeftijd ( <i>age</i> )                             | 1   | 0.00 |      | 0.00  | 0.00 | 0.00 |

*Continued on next page*

| Dataset                                    | Tag                                    | N    | Det. | rec. | Prec. | Rec. | F1   |
|--------------------------------------------|----------------------------------------|------|------|------|-------|------|------|
|                                            | persoon ( <i>person</i> )              | 66   | 0.92 |      | 0.86  | 0.86 | 0.86 |
|                                            | persoonafkorting ( <i>initials</i> )   | 2    | 1.00 |      | 0.50  | 1.00 | 0.67 |
|                                            | telefoonnummer ( <i>phone no.</i> )    | 14   | 1.00 |      | 0.86  | 0.86 | 0.86 |
|                                            | tijd ( <i>time</i> )                   | 30   | 0.97 |      | 0.97  | 0.93 | 0.95 |
|                                            | url ( <i>URL</i> )                     | 1    | 1.00 |      | 1.00  | 1.00 | 1.00 |
|                                            | ziekenhuis ( <i>hospital</i> )         | 8    | 0.75 |      | 0.45  | 0.62 | 0.53 |
| <b>ZGT</b> ( $N = 239$ )                   |                                        |      |      |      |       |      |      |
|                                            | adres ( <i>address</i> )               | 7    | 1.00 |      | 0.62  | 0.71 | 0.67 |
|                                            | agbnummer ( <i>AGB no.</i> )           | 42   | 1.00 |      | 0.98  | 1.00 | 0.99 |
|                                            | bedrijf ( <i>company</i> )             | 2    | 0.50 |      | 0.20  | 0.50 | 0.29 |
|                                            | bignummer ( <i>BIG no.</i> )           | 35   | 1.00 |      | 1.00  | 1.00 | 1.00 |
|                                            | bsn ( <i>citizen no.</i> )             | 7    | 1.00 |      | 1.00  | 1.00 | 1.00 |
|                                            | datum ( <i>date</i> )                  | 1264 | 0.99 |      | 0.97  | 0.97 | 0.97 |
|                                            | documentid ( <i>document ID</i> )      | 245  | 1.00 |      | 0.99  | 1.00 | 0.99 |
|                                            | documentnummer ( <i>document no.</i> ) | 252  | 1.00 |      | 0.99  | 0.99 | 0.99 |
|                                            | leeftijd ( <i>age</i> )                | 71   | 0.93 |      | 0.77  | 0.93 | 0.84 |
|                                            | patientnummer ( <i>patient no.</i> )   | 11   | 0.91 |      | 0.90  | 0.82 | 0.86 |
|                                            | persoon ( <i>person</i> )              | 427  | 0.97 |      | 0.93  | 0.86 | 0.90 |
|                                            | phinummer ( <i>other PHI no.</i> )     | 31   | 0.97 |      | 0.75  | 0.68 | 0.71 |
|                                            | plaats ( <i>place</i> )                | 13   | 0.85 |      | 0.60  | 0.69 | 0.64 |
|                                            | studie naam ( <i>study name</i> )      | 8    | 1.00 |      | 0.60  | 0.75 | 0.67 |
|                                            | telefoonnummer ( <i>phone no.</i> )    | 14   | 1.00 |      | 0.81  | 0.93 | 0.87 |
|                                            | tijd ( <i>time</i> )                   | 96   | 0.91 |      | 0.89  | 0.89 | 0.89 |
|                                            | ziekenhuis ( <i>hospital</i> )         | 150  | 0.92 |      | 0.75  | 0.79 | 0.77 |
| <b>JBZ external test set</b> ( $N = 340$ ) |                                        |      |      |      |       |      |      |

*Continued on next page*

| Dataset | Tag                                                 | N   | Det. rec. | Prec. | Rec. | F1   |
|---------|-----------------------------------------------------|-----|-----------|-------|------|------|
|         | accreditatie<br>nummer ( <i>accreditation no.</i> ) | 100 | 1.00      | 1.00  | 1.00 | 1.00 |
|         | bignummer ( <i>BIG no.</i> )                        | 3   | 1.00      | 0.04  | 0.33 | 0.08 |
|         | datum ( <i>date</i> )                               | 807 | 0.99      | 0.99  | 0.99 | 0.99 |
|         | leeftijd ( <i>age</i> )                             | 78  | 0.99      | 0.94  | 0.99 | 0.96 |
|         | patientnummer ( <i>patient no.</i> )                | 70  | 1.00      | 0.97  | 0.47 | 0.63 |
|         | persoon ( <i>person</i> )                           | 482 | 0.97      | 0.98  | 0.95 | 0.96 |
|         | persoonafkorting ( <i>initials</i> )                | 77  | 1.00      | 0.94  | 0.99 | 0.96 |
|         | phinummer ( <i>other PHI no.</i> )                  | 2   | 1.00      | 0.00  | 0.00 | 0.00 |
|         | plaats ( <i>place</i> )                             | 5   | 0.80      | 0.67  | 0.80 | 0.73 |
|         | rapport id.c nummer ( <i>report ID C-no.</i> )      | 41  | 1.00      | 1.00  | 1.00 | 1.00 |
|         | rapport id.t nummer ( <i>report ID T-no.</i> )      | 62  | 1.00      | 1.00  | 1.00 | 1.00 |
|         | telefoonnummer ( <i>phone no.</i> )                 | 21  | 1.00      | 0.94  | 0.71 | 0.81 |
|         | tijd ( <i>time</i> )                                | 58  | 0.97      | 0.97  | 0.97 | 0.97 |
|         | ziekenhuis ( <i>hospital</i> )                      | 12  | 1.00      | 0.48  | 0.92 | 0.63 |

### S3. False-Positive Analysis: Inadvertent Information Loss

Table S3 summarises how predicted PHI spans relate to gold annotations across both test sets. A span is classified as an exact true positive if span boundaries and tag match exactly, an overlap error if it overlaps any gold span but with incorrect boundaries or tag, and a pure false positive if it has no token-level overlap with any gold span. Table S4 breaks down pure false positives by predicted tag.

Table S3: Predicted span classification by test set. Exact TP: span boundaries and tag match exactly. Overlap: overlaps a gold span but with wrong boundaries or tag type. Pure FP: no token-level overlap with any gold span; these represent clinical content the pipeline would replace with a surrogate despite not being PHI. Percentages are of total predicted spans for that row.

| Subset                | Cases        | Pred.        | Exact TP             | Overlap           | Pure FP           | FP/case     |
|-----------------------|--------------|--------------|----------------------|-------------------|-------------------|-------------|
| Internal (RUMC & ZGT) | 980          | 4,911        | 4,623 (94.1%)        | 197 (4.0%)        | 91 (1.9%)         | 0.09        |
| External (JBZ)        | 340          | 1,823        | 1,730 (94.9%)        | 74 (4.1%)         | 19 (1.0%)         | 0.06        |
| <b>Combined</b>       | <b>1,320</b> | <b>6,734</b> | <b>6,353 (94.3%)</b> | <b>271 (4.0%)</b> | <b>110 (1.6%)</b> | <b>0.08</b> |

Table S4: Pure false positives by predicted tag across both test sets combined. Tags not appearing in this table had zero pure false positives. FP rate is the fraction of all predictions for that tag that were pure false positives.

| Tag                                  | Pure FP    | Total predicted | FP rate     |
|--------------------------------------|------------|-----------------|-------------|
| ziekenhuis ( <i>hospital</i> )       | 27         | 306             | 8.8%        |
| leeftijd ( <i>age</i> )              | 23         | 172             | 13.4%       |
| persoonafkorting ( <i>initials</i> ) | 19         | 517             | 3.7%        |
| datum ( <i>date</i> )                | 13         | 3,192           | 0.4%        |
| persoon ( <i>person</i> )            | 13         | 1,031           | 1.3%        |
| studie“ naam ( <i>study name</i> )   | 4          | 14              | 28.6%       |
| plaats ( <i>place</i> )              | 3          | 22              | 13.6%       |
| iban ( <i>IBAN</i> )                 | 3          | 3               | 100.0%      |
| bsn ( <i>citizen no.</i> )           | 2          | 9               | 22.2%       |
| bedrijf ( <i>company</i> )           | 1          | 6               | 16.7%       |
| tijd ( <i>time</i> )                 | 1          | 186             | 0.5%        |
| rapport“ id ( <i>report ID</i> )     | 1          | 198             | 0.5%        |
| <b>Total</b>                         | <b>110</b> | <b>6,734</b>    | <b>1.6%</b> |

## S4. Baseline Performances per Dataset

This section provides the full per-subset breakdown of detection recall for GHMSCHRFT, DEDUCE, RRA, and the OpenAI Privacy Filter across compiled PHI categories. The main text reports only the aggregate internal and external results; this table shows the individual contributions of each dataset subset. Note that RRA was built and tuned specifically for Radboudumc radiology reports, which is reflected in its strong performance on that subset relative to ZGT and JBZ.

Table S5: Detection recall comparison between GHMSCHRFT, two rule-based baselines, and the OpenAI Privacy Filter, broken down by dataset subset and compiled PHI category. The **Overall** rows for the combined internal and external test sets report support-weighted recall with 95% bootstrap confidence intervals (5,000 iterations, document-level resampling) in parentheses. PF = OpenAI Privacy Filter.

| Category                            | N           | GHMSCHRFT    | DEDUCE       | RRA          | PF           |
|-------------------------------------|-------------|--------------|--------------|--------------|--------------|
| <b>RUMC radiology</b> ( $N = 667$ ) |             |              |              |              |              |
| persoon ( <i>person</i> )           | 68          | 0.93         | 0.72         | 0.78         | 0.52         |
| datum ( <i>date</i> )               | 571         | 1.00         | 0.78         | 0.97         | 0.63         |
| leeftijd ( <i>age</i> )             | 1           | 0.00         | 0.00         | 0.00         | 0.00         |
| telefoonnummer ( <i>phone no.</i> ) | 14          | 1.00         | 0.29         | 0.57         | 0.50         |
| url ( <i>URL</i> )                  | 1           | 1.00         | 1.00         | 0.00         | 0.00         |
| locatie ( <i>location</i> )         | 4           | 0.75         | 0.25         | 0.00         | 0.00         |
| ziekenhuis ( <i>hospital</i> )      | 8           | 0.75         | 0.25         | 0.13         | 0.25         |
| <b>Overall</b>                      | <b>667</b>  | <b>0.983</b> | <b>0.758</b> | <b>0.919</b> | <b>0.604</b> |
| <b>RUMC pathology</b> ( $N = 199$ ) |             |              |              |              |              |
| persoon ( <i>person</i> )           | 526         | 0.98         | 0.19         | 0.19         | 0.13         |
| datum ( <i>date</i> )               | 571         | 1.00         | 0.97         | 0.61         | 0.36         |
| url ( <i>URL</i> )                  | 7           | 1.00         | 0.86         | 0.00         | 0.00         |
| ziekenhuis ( <i>hospital</i> )      | 107         | 0.99         | 0.46         | 0.23         | 0.01         |
| id ( <i>identifier</i> )            | 313         | 0.99         | 0.00         | 0.28         | 0.05         |
| <b>Overall</b>                      | <b>1524</b> | <b>0.991</b> | <b>0.467</b> | <b>0.367</b> | <b>0.192</b> |
| <b>ZGT</b> ( $N = 239$ )            |             |              |              |              |              |

*Continued on next page*

| Category                                                                 | N           | GHMSCHRFT    | DEDUCE       | RRA          | PF           |
|--------------------------------------------------------------------------|-------------|--------------|--------------|--------------|--------------|
| persoon ( <i>person</i> )                                                | 427         | 0.97         | 0.71         | 0.24         | 0.43         |
| datum ( <i>date</i> )                                                    | 1264        | 0.99         | 0.49         | 0.39         | 0.31         |
| leeftijd ( <i>age</i> )                                                  | 71          | 0.93         | 0.92         | 0.03         | 0.00         |
| telefoonnummer ( <i>phone no.</i> )                                      | 14          | 1.00         | 0.50         | 0.07         | 0.86         |
| locatie ( <i>location</i> )                                              | 20          | 0.90         | 0.55         | 0.30         | 0.30         |
| ziekenhuis ( <i>hospital</i> )                                           | 150         | 0.92         | 0.34         | 0.15         | 0.05         |
| id ( <i>identifier</i> )                                                 | 623         | 1.00         | 0.56         | 0.07         | 0.51         |
| <b>Overall</b>                                                           | <b>2569</b> | <b>0.982</b> | <b>0.545</b> | <b>0.258</b> | <b>0.357</b> |
| <b>Combined internal test set (RUMC &amp; ZGT; <math>N = 980</math>)</b> |             |              |              |              |              |
| persoon ( <i>person</i> )                                                | 1021        | 0.97         | 0.44         | 0.25         | 0.28         |
| datum ( <i>date</i> )                                                    | 2406        | 1.00         | 0.67         | 0.58         | 0.40         |
| leeftijd ( <i>age</i> )                                                  | 72          | 0.92         | 0.90         | 0.03         | 0.00         |
| telefoonnummer ( <i>phone no.</i> )                                      | 28          | 1.00         | 0.39         | 0.32         | 0.68         |
| locatie ( <i>location</i> )                                              | 24          | 0.88         | 0.50         | 0.25         | 0.25         |
| ziekenhuis ( <i>hospital</i> )                                           | 265         | 0.94         | 0.39         | 0.18         | 0.04         |
| id ( <i>identifier</i> )                                                 | 936         | 0.99         | 0.37         | 0.14         | 0.36         |
| <b>Overall</b>                                                           | <b>4752</b> | <b>0.985</b> | <b>0.549</b> | <b>0.386</b> | <b>0.339</b> |
|                                                                          |             | (0.98, 0.99) | (0.53, 0.57) | (0.36, 0.41) | (0.31, 0.36) |
| <b>External test set (JBZ; <math>N = 340</math>)</b>                     |             |              |              |              |              |
| persoon ( <i>person</i> )                                                | 559         | 0.98         | 0.70         | 0.11         | 0.46         |
| datum ( <i>date</i> )                                                    | 807         | 0.99         | 0.71         | 0.44         | 0.62         |
| leeftijd ( <i>age</i> )                                                  | 78          | 0.99         | 0.95         | 0.04         | 0.00         |
| telefoonnummer ( <i>phone no.</i> )                                      | 21          | 1.00         | 0.43         | 0.00         | 0.05         |
| locatie ( <i>location</i> )                                              | 5           | 0.80         | 0.40         | 0.60         | 0.00         |
| ziekenhuis ( <i>hospital</i> )                                           | 12          | 1.00         | 0.33         | 0.00         | 0.00         |
| id ( <i>identifier</i> )                                                 | 175         | 1.00         | 0.42         | 0.38         | 0.38         |
| <b>Overall</b>                                                           | <b>1657</b> | <b>0.986</b> | <b>0.682</b> | <b>0.295</b> | <b>0.498</b> |
|                                                                          |             | (0.98, 0.99) | (0.65, 0.72) | (0.27, 0.32) | (0.46, 0.54) |

## S5. Nemotron Privacy Filter: Supplementary Evaluation

The Nemotron Privacy Filter [?] is a 1.4B sparse mixture-of-experts model fine-tuned on English medical text for PHI detection across 55 categories. It

uses BIOES span encoding and was included as a supplementary comparator to test whether fine-tuning on English medical text improves over the out-of-the-box OpenAI Privacy Filter on Dutch clinical data. Because it is an English-medical fine-tuned variant of the same model family, this comparison isolates the effect of English medical fine-tuning.

Table S6 shows compiled-category detection recall for both the OpenAI Privacy Filter (PF) and the Nemotron Privacy Filter on the internal and external test sets, with 95% bootstrap confidence intervals (5,000 iterations, document-level resampling).

Table S6: Compiled-category detection recall for the OpenAI Privacy Filter (PF) and the Nemotron Privacy Filter (Nemotron) on the internal (RUMC & ZGT;  $N = 980$ ) and external (JBZ;  $N = 340$ ) test sets.

| Category                       | Internal test set |              |              | External test set (JBZ) |              |              |
|--------------------------------|-------------------|--------------|--------------|-------------------------|--------------|--------------|
|                                | N                 | PF           | Nemotron     | N                       | PF           | Nemotron     |
| persoon ( <i>person</i> )      | 1021              | 0.28         | 0.26         | 559                     | 0.46         | 0.42         |
| datum ( <i>date</i> )          | 2406              | 0.40         | 0.44         | 807                     | 0.62         | 0.66         |
| leeftijd ( <i>age</i> )        | 72                | 0.00         | 0.07         | 78                      | 0.00         | 0.01         |
| telefoonnummer ( <i>tel.</i> ) | 28                | 0.68         | 0.54         | 21                      | 0.05         | 0.33         |
| locatie ( <i>location</i> )    | 24                | 0.25         | 0.33         | 5                       | 0.00         | 0.60         |
| ziekenhuis ( <i>hospital</i> ) | 265               | 0.04         | 0.13         | 12                      | 0.00         | 0.00         |
| id ( <i>identifier</i> )       | 936               | 0.36         | 0.13         | 175                     | 0.38         | 0.03         |
| <b>Overall</b>                 | <b>4752</b>       | <b>0.339</b> | <b>0.315</b> | <b>1657</b>             | <b>0.498</b> | <b>0.474</b> |
|                                |                   | (0.31, 0.36) | (0.30, 0.33) |                         | (0.46, 0.54) | (0.45, 0.50) |

The Nemotron variant did not outperform the base OpenAI Privacy Filter on overall recall (0.315 vs. 0.339 internal; 0.474 vs. 0.498 external). The most notable divergence appeared in the identifier category, where Nemotron recalled substantially fewer spans than the base model (0.126 vs. 0.359 internal; 0.034 vs. 0.377 external). This is consistent with identifiers in the Dutch clinical context following Netherlands-specific conventions (e.g., AGB numbers, RUMC report IDs) that are not represented in English medical fine-tuning data. Nemotron showed modestly higher recall for hospital names (0.128 vs.

0.038 internal) and age expressions (0.069 vs. 0.000 internal), both of which are categories where the base model performs near zero. Overall, English medical fine-tuning does not compensate for the lack of Dutch language coverage.

## S6. Compiled Categories for Baseline Comparison

Table S7 lists the eight compiled PHI categories used for detection-recall comparison across GHMSCHRFT, DEDUCE, RRA, and the OpenAI Privacy Filter. Categories were selected to have clear counterparts across all four systems; GHMSCHRFT-only categories without a direct counterpart in any baseline were excluded.

Table S7: Compiled categories used for detection-recall comparison across GHMSCHRFT, DEDUCE, RRA, and the OpenAI Privacy Filter. GHMSCHRFT-only categories without a direct counterpart in any baseline were excluded. For the external (JBZ) test set, the <RAPPORT\ ID> subtypes <RAPPORT\ ID.C\ NUMMER> and <RAPPORT\ ID.T\ NUMMER> were additionally excluded from the ID compiled category, as these JBZ-specific formats lack counterparts in any baseline system.

| Compiled category                   | GHMSCHRFT tags included                                                                                                                    |
|-------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------|
| persoon ( <i>person</i> )           | <PERSOON>, <PERSOONAFKORTING>                                                                                                              |
| datum ( <i>date</i> )               | <DATUM>                                                                                                                                    |
| leeftijd ( <i>age</i> )             | <LEEFTIJD>                                                                                                                                 |
| telefoonnummer ( <i>phone no.</i> ) | <TELEFOONNUMMER>                                                                                                                           |
| locatie ( <i>location</i> )         | <PLAATS>, <ADRES>                                                                                                                          |
| ziekenhuis ( <i>hospital</i> )      | <ZIEKENHUIS>                                                                                                                               |
| id ( <i>identifier</i> )            | <BSN>, <PATIENTNUMMER>, <BIGNUMMER><br><AGBNUMMER>, <RAPPORT\ ID>, <DOCUMENTID><br><DOCUMENTNUMMER>, <ACCREDITATIE\ NUMMER><br><PHINUMMER> |

## S7. PHI Annotation Guidelines

This appendix defines the PHI labeling policy used for all manual annotations in this study. The guidelines were developed iteratively during the annotation process and finalized before the main annotation rounds began. They cover 24 PHI categories and include span-merging rules, priority rules where categories overlap, and explicit edge-case examples. Annotators were trained on these guidelines; inter-reader agreement was assessed on dedicated doubly annotated subsets (see Section 2.1 of the main text).

**General tagging rule.** When multiple consecutive PHI references point to the same person or place, they are annotated as a single span.

### PERSOON (person)

**Includes:** first names, last names, and initials; consecutive combinations are labeled as one span.

**Excludes:** eponymous disease or body-part names (e.g. *Creutzfeldt-Jakob*); report-level specialty fields (e.g. *Specialisme: Urologie*).

*Informatie doorgebeld aan radioloog A. Arends*

→ *Informatie doorgebeld aan radioloog* PERSOON

*In behandeling bij Stephan*

→ *In behandeling bij* PERSOON

*gebeld met collega Breedijk (hematologie)*

→ *gebeld met collega* PERSOON *(hematologie)*

## PERSOONAFKORTING (abbreviated person reference)

**Includes:** abbreviations and titled shorthand references to persons.

*Ingepakt door LPat*

→ *Ingepakt door* PERSOONAFKORTING

*T.i. 1b. RRov/KZoe*

→ *T.i. 1b.* PERSOONAFKORTING / PERSOONAFKORTING

*verslag geautoriseerd door AEij, KMBP*

→ *verslag geautoriseerd door* PERSOONAFKORTING, PERSOONAFKORTING

## ZIEKENHUIS (hospital)

**Includes:** explicit hospital names and place names that implicitly refer to a hospital.

**Excludes:** department names.

*Overgeplaatst naar Radboud Nijmegen*

→ *Overgeplaatst naar* ZIEKENHUIS

*Verwijzing vanuit AMC*

→ *Verwijzing vanuit* ZIEKENHUIS

*Doorverwezen vanuit Leiden zkhs*

→ *Doorverwezen vanuit* ZIEKENHUIS

*Beelden uit Nijmegen*

→ *Beelden uit* ZIEKENHUIS

*Genoomdiagnostiek Nijmegen*

→ *Genoomdiagnostiek* ZIEKENHUIS

*Laboratorium Medische Microbiologie* → *<not tagged>*

## PLAATS (place)

**Includes:** place names that do not refer to a hospital.

*Woont in Almere*

→ *Woont in* PLAATS

*Persoon uit België*

→ *Persoon uit* PLAATS

## ADRES (address)

**Includes:** all full addresses, including street names, house numbers, postal codes, and city names; multiline addresses are collapsed into a single span.

*Woonadres: Kerkstraat 12, 1234 AB Amsterdam*

→ *Woonadres:* ADRES

*Multiline:*

*Adres:* → *Adres:*

*Kerkstraat 12*

*1234 AB Amsterdam*

ADRES

## BEDRIJF (company / employer)

**Includes:** organization names only when indicating the employer or physical workplace of a specific person.

**Excludes:** manufacturers, vendors, suppliers, test-kit producers, or organizations not linked to a specific person.

*Werkzaam bij Sanquin*

→ *Werkzaam bij* BEDRIJF

*Patholoog in dienst van Roche*

→ *Patholoog in dienst van* BEDRIJF

## DATUM (date)

**Includes:** all date formats.

*Gemeten op 28/03*

→ *Gemeten op* DATUM

*Laatste scan in 2011*

→ *Laatste scan in* DATUM

*Vergeleken werd met scans van 13 en 29 juli*

→ *Vergeleken werd met scans van* DATUM *en* DATUM

## TIJD (time)

**Includes:** all time formats.

**Excludes:** anatomical clock positions.

*Opname om 18:03*

→ *Opname om* TIJD

*Scan genomen rond 20u*

→ *Scan genomen rond* TIJD

*Laesie op 6 uur* → *<not tagged>*

## LEEFTIJD (age)

**Includes:** patient age expressions.

*Patient is 49 jaar oud*

→ *Patient is* LEEFTIJD *jaar oud*

*Leeftijd: 86 jaar*

→ *Leeftijd:* LEEFTIJD *jaar*

#### TELEFOONNUMMER (phone number)

**Includes:** internal and external phone numbers.

*Arts gebeld op sein 6162*

→ *Arts gebeld op* TELEFOONNUMMER

*Doorgebeld op 45685*

→ *Doorgebeld op* TELEFOONNUMMER

*Bereikbaar op 06-15468560*

→ *Bereikbaar op* TELEFOONNUMMER

#### EMAIL (email address)

**Includes:** internal and external email addresses.

*Contact via jansen@radboudumc.nl*

→ *Contact via* EMAIL

*Mail gestuurd naar info@labservice.com*

→ *Mail gestuurd naar* EMAIL

#### URL (URL)

**Includes:** all URLs.

*<https://www.palga.nl/datasheet/Radboudumc/TS0500.pdf>*

→ URL

#### PATIENTNUMMER (patient number)

**Includes:** Radboudumc patient IDs consisting of exactly seven digits.

*Patientnummer 1597557*

→ *Patientnummer* PATIENTNUMMER

#### ZNUMMER (Z-number)

**Includes:** Radboudumc personnel Z-numbers.

*Geholpen door Z854913*

→ *Geholpen door* ZNUMMER

#### RAPPORT\_ID (report identifier)

**Includes:** Radboudumc report identifiers; the subtype is indicated by a dot-suffix label.

T12-38295 → RAPPORT\_ID.T.NUMMER

R 04-167190 → RAPPORT\_ID.R.NUMMER

C22-6116 → RAPPORT\_ID.C.NUMMER

DPA18-00935\_23849971 → RAPPORT\_ID.DPA.NUMMER

RPA22-00473 → RAPPORT\_ID.RPA.NUMMER

#### DOCUMENTID (document identifier)

**Includes:** system-level document identifiers assigned by the hospital's document management system to uniquely identify a clinical document.

*DocumentID: 348*

→ *DOCUMENTID* DOCUMENTID

#### DOCUMENTNUMMER (document number)

**Includes:** human-readable document sequence numbers used in clinical correspondence to identify the version or instance of a document.

*Documentnummer: 20210445*

→ *DOCUMENTNUMMER* DOCUMENTNUMMER

#### PHINUMMER (other PHI number)

**Includes:** PHI numbers not covered by `PATIENTNUMBER`, `ZNUMBER`, `RAPPORT_ID`, `DOCUMENTID`, or `DOCUMENTNUMBER` (e.g., external IDs, order numbers).

*Patient 64852145*

→ *Patient* `PHINUMMER`

*Zie order 160637831*

→ *Zie order* `PHINUMMER`

#### ACCREDITATIE\_NUMMER (accreditation number)

**Includes:** accreditation numbers of health institutes.

*onder ISO 15189 accreditatie M135*

→ `ACCREDITATIE_NUMMER`

*Accreditatie nummer M142*

→ `ACCREDITATIE_NUMMER`

#### IBAN (IBAN)

**Includes:** IBAN numbers from any country.

*IBAN: NL91ABNA0417164300*

→ *IBAN:* `IBAN`

*Rekeningnummer BE68 5390 0754 7034*

→ *Rekeningnummer* `IBAN`

#### BSN (citizen service number)

**Includes:** all Dutch BSN numbers.

**Priority rule:** BSN values are always tagged as `BSN`, never as `PHINUMMER`.

*BSN 123456782*

→ BSN

*BSN: 228956781*

→ BSN

#### BIGNUMBER (BIG registration number)

**Includes:** BIG registration numbers of healthcare professionals.

*BIG-nummer 19012345678*

→ BIGNUMMER

*BIG registratie: 29098765432*

→ BIGNUMMER

#### AGBNUMMER (AGB code)

**Includes:** AGB codes of healthcare providers or institutions.

*AGB-code 03001234*

→ AGBNUMMER

*AGB nummer: 08009999*

→ AGBNUMMER

#### STUDIE\_NAAM (study name)

**Includes:** study names and study codes (e.g. LEMA, PINNACLE, DONAN, MSPECT).

*LEMA studiecode UZR0291*

→ *STUDIE\_NAAM*

*LEMA studie*

→ *STUDIE\_NAAM*

*LEMA*

→ *STUDIE\_NAAM*

*donan 108329*

→ *STUDIE\_NAAM*