Quality evaluation and analysis of linked data sets

Training course on record linkage and statistical matching

*Tiziana Tuoto*

*Istat*

tuoto@istat.it
Quality evaluation

Training course on record linkage and statistical matching

Tiziana Tuoto
Istat
tuoto@istat.it
Outline

- Linkage errors
- Formal definition of linkage errors
- Methods for linkage errors evaluation
  - Mixture Models approaches
  - Sampling and clerical resolution
  - Other approaches
- Prevent linkage errors
Linkage errors

- *We know that few pairs in the Linked set are not in fact true pairs*

- *We know that some pairs in the grey set of Possible Links are not in fact true pairs*

- *We know that most pairs in the Unlinked set are not in fact true pairs*
Linkage errors

Two kinds of errors:

- **false matches:** the pair is declared as a link but actually the two records are referred to different units

- **missed matches:** the pair is declared as a non-link but actually the two records are referred to the same units
Linkage errors

*In statistics, linkage errors are referred as false match rate and false non-match rate.*

*In other contexts (medicine, epidemiology), they are known as positive predicted value and sensitivity.*

*In information technology they are named precision and recall.*
Linkage errors

The RL strategies often need to reach trade-offs between false matches and missed matches: decreasing the number of false positives (the false matches) will increase the number of false negatives (the missed matches)
The two errors can have different importance depending on the objectives of the specific linkage problem.

Generally, the missed matches can be considered less serious of the false matches.
Linkage errors

Example 1. Objective: joint analysis of variables stored in distinct files. The false matches have to be avoided, since they could introduce bias in the standard estimates of the statistic relationship among the variables.
Linkage errors

Example 2. Objective: creation of a list to be used as sample frame. In this case, it could be preferable to include as much matches as possible in order to avoid possible under-count of the list. The false matches cause the inclusion of non-eligible units in the list (these can be deleted in a following steps), but the more serious risk is the exclusion of eligible units (missed matches).
Linkage errors

Further criteria for the evaluation of the performances of linkage procedures are the time complexity of the software and the amount of records that need manual revision.
Linkage errors

The time complexity of the linkage algorithm is generally dominated by the number of comparison among records.

The manual revision is a delicate activity that needs time, resources and is error-prone.
## Formal definition of linkage errors

<table>
<thead>
<tr>
<th>Results of the Linkage Procedures</th>
<th>True Linkage Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matches</td>
<td>Non-Matches</td>
</tr>
<tr>
<td>Linked</td>
<td></td>
</tr>
<tr>
<td>Linked</td>
<td>$a$ (true positive)</td>
</tr>
<tr>
<td>Non-Linked</td>
<td></td>
</tr>
<tr>
<td>Non-Linked</td>
<td>$b$ (false positive or false links)</td>
</tr>
<tr>
<td>Non-Linked</td>
<td>$c$ (false negative or missed links)</td>
</tr>
<tr>
<td>Non-Linked</td>
<td>$d$ (true negative)</td>
</tr>
</tbody>
</table>
Formal definition of linkage errors

<table>
<thead>
<tr>
<th>Results of the Linkage Procedures</th>
<th>True Linkage Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linked</td>
<td>Matches</td>
</tr>
<tr>
<td></td>
<td>Non-Matches</td>
</tr>
<tr>
<td>Linked</td>
<td>(a)</td>
</tr>
<tr>
<td>Non-Linked</td>
<td>(b)</td>
</tr>
</tbody>
</table>

\[
\lambda = \frac{c}{a + c} \\
\mu = \frac{b}{b + d}
\]

- **False non-match rate:** \(\lambda\)  
  \((= 1 - \text{sensitivity} = 1 - \text{recall})\)

- **False match rate:** \(\mu\)  
  \((= 1 - \text{specificity})\)
Formal definition of linkage errors

- **False match rate:**
  \[ \frac{b}{a + b} \]
  \[ (=1 - \text{precision} = 1 - \text{match accuracy}) \]

- **Analogously:**
  \[ \frac{c}{c + d} \]
Methods for errors estimation

- Mixture Models approaches
  - Fellegi-Sunter
  - Belin-Rubin
- Sampling-based Clerical Resolution
- Other Approaches and Heuristics
  - Methods based on probabilities of any chance links
  - Comparison with a benchmark file (Gold-Standard)
  - Counting number of duplicate links (in 1:1 linkage)
Mixture Models approach: the Fellegi-Sunter estimation

According to the Fellegi and Sunter model, the pair classification criterion is based on the two thresholds: $T_m$ and $T_u$ ($T_m > T_u$)

$$r_{(a,b)} > T_m \implies (a,b) \in M^*$$

$$T_m \geq r_{(a,b)} \geq T_u \implies (a,b) \in Q$$

$$r_{(a,b)} < T_u \implies (a,b) \in U^*$$
The Fellegi-Sunter errors estimation

The Fellegi and Sunter decision rule is based on the error levels:

*a pair* \((a, b) \in U\) *can in fact assume a value* \(r > T_m\)

**The frequency of the error due to the decision** \(M^*\) **is**

\[
\mu = \sum_{\gamma \in \Gamma} u(\gamma) P(M^* | \gamma) = \sum_{\gamma \in \Gamma_{M^*}} u(\gamma) \quad \text{where} \quad \Gamma_{M^*} = \left\{ \gamma : T_m \leq m(\gamma) / u(\gamma) \right\}
\]
In the same way, a pair \((a, b) \in M\) can in fact assume a value \(r < T_u\).

The frequency of the error due to the decision \(U^*\) is

\[
\lambda = \sum_{\gamma \in \Gamma} m(\gamma) P(U^* | \gamma) = \sum_{\gamma \in \Gamma_{U^*}} m(\gamma) \quad \text{where} \quad \Gamma_{U^*} = \{ \gamma : T_u \geq m(\gamma) / u(\gamma) \}
\]
The Fellegi-Sunter errors estimation

\[ r = \frac{m(\gamma)}{u(\gamma)} \]

Density

Weight values

Error \( \lambda \)

\( T_u \)  \( T_m \)

Error \( \mu \)

\( U^* \)

\( M^* \)
The Fellegi-Sunter errors estimation

Fellegi and Sunter suggest to select “acceptable” values for the error levels \( \mu \) and \( \lambda \). Fixed these error values, the thresholds \( T_m \) and \( T_u \) can be obtained solving the previous formula.

After \( m(\gamma) \) and \( u(\gamma) \) distributions estimation, \( \mu \) and \( \lambda \) can be estimated as well.

\[
\hat{\mu} = \sum_{\gamma \in \Gamma^*_M} \hat{u}(\gamma) \quad \quad \hat{\lambda} = \sum_{\gamma \in \Gamma^*_U} \hat{m}(\gamma)
\]
The Fellegi-Sunter estimation: comments

- Strong dependence on the accuracy of the estimates of $m(\gamma)$ and $u(\gamma)$.
- Errors in the model assumption specification (e.g. the reliability of the conditional independence assumption), lack of information and so on can cause loss in estimates accuracy and consequently a strong bias in the estimates of the errors.
The Fellegi-Sunter estimation: comments

- Model assumptions violations could have a strong effect on the linkage error estimation;

- They usually however have only a minor effect on the quality on the matching because the only purpose of the scores is to achieve a correct rank order of the record pairs.
Mixture Models approach: the Belin and Rubin method

Belin and Rubin, JASA 1995

Aim: evaluation of the false match rates for each possible threshold value

Objective: the probability to be a true link with the relative standard error (expressed as function of the linkage weights) is associated to each pair
Mixture Models approach: the Belin and Rubin method

Requirements: a training sample of pairs where the true linkage status is known

Assumption of the model: the linkage weights distribution is a mixture of the true link distribution and the true non-link one.
Mixture Models approach: the Belin and Rubin method

The method works well when:

- the linkage is 1:1;
- there is a good separation between linkage weights associated to matches and non-matches;
- slight failure of the conditional independence assumption is also allowed.
Mixture Models approach: the Belin and Rubin method

1. Start point: mixture model of two Normal distributions, with unknown means and variances (throughout Box-Cox transformations)

2. Estimate the power of the two transformations (the shapes of the curves) from the training set
Mixture Models approach: the Belin and Rubin method

3. Considering known the transformations, evaluate MLE and SE of the normals mixture model

4. Throughout the model estimates, obtain punctual estimates for a) false match rate (as function of the thresholds), b) SE of the false match rate (with delta approximation)
The Belin and Rubin methods: comments

- Limited only to the estimation of the false match rate (false positive)
- Requires the use of its own sample of training data to obtain parameter estimates
- While the method is robust to independence assumptions and asymptotic SEs can be computed, distributional assumptions may not be valid
Sampling and clerical resolution

Estimates of errors rates can be made through sampling pairs from the linkage procedure and subjecting the pairs to clerical resolution. Estimates of the quality indicators can be derived from the results of the clerical resolution. This is actually the most frequent method.
Sampling and clerical resolution

1. Select a sample from the set of the linkage candidate pairs, both from pairs assigned to the set $M^*$ and from pairs assigned to the set $U^*$.

Sampling can be random or judgmental.
Sampling and clerical resolution

2. The sampled pairs are re-linked by manual review from very expert staff.

3. The “rematch” is considered as “perfect”, i.e. error free and representative of the true linkage status.
Sampling and clerical resolution

4. The potential bias of the original linkage is evaluated throughout the discrepancies among the results of the first linkage and the “rematch”.

5. Linkage errors can be “extended” to the whole set of pairs
## Sampling and clerical resolution

### True Linkage Status (Results of Re-match)

<table>
<thead>
<tr>
<th>Results of the Linkage Procedures</th>
<th>Matched</th>
<th>Not Matched</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matched</td>
<td>$n_{11}$</td>
<td>$n_{12}$</td>
<td>$n_{1.}$</td>
</tr>
<tr>
<td>Not Matched</td>
<td>$n_{21}$</td>
<td>$n_{22}$</td>
<td>$n_{2.}$</td>
</tr>
<tr>
<td>Total</td>
<td>$n_{.1}$</td>
<td>$n_{.2}$</td>
<td></td>
</tr>
</tbody>
</table>
# Sampling and clerical resolution

<table>
<thead>
<tr>
<th>Results of the Linkage Procedures</th>
<th>True Linkage Status (Results of Re-match)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Matched</td>
</tr>
<tr>
<td>Linked</td>
<td>( n_{11} )</td>
</tr>
<tr>
<td>Non-Linked</td>
<td>( n_{21} )</td>
</tr>
<tr>
<td>Total</td>
<td>( n_{.1} )</td>
</tr>
</tbody>
</table>

- **False Positive** \( n_{12} \)
- **False Negative** \( n_{21} \)
Sampling and clerical resolution: comments

- how to select the pairs to include in the sample?
- how much? for instance, searching for false negative among non-linked pairs is frequently compared to searching for a needle in a haystack...
- what is the cost?
- who should do it?
Sampling and clerical resolution: comments

Recognizing that resources may well be limited:

- Winkler (1995) suggest to reduce the sample size throughout the selection of the pairs in the grey area, near to the thresholds, by means a weighted sample strategy;
Sampling and clerical resolution: comments

Recognizing that resources may well be limited:

- optimal stratified sampling strategy required to achieve a reasonable accuracy estimated is explored in Heasman (2014).
Other Approaches for linkage errors estimation

- Methods based on probabilities of any chance links

As the probability of any chance matches decrease, confidence in any identified links increase. False matches rate is estimated as the ratio between the estimated number of chance matches and the total number of achieved links.
Methods based on probabilities of any chance links

Assumptions on the matching variables distributions are needed

Data are only used to estimate distributional parameters or verify assumptions

No additional record linkage is needed

Other Approaches for linkage errors estimation

- Comparison with a benchmark file (Gold-Standard)
- Counting number of duplicate links (in 1:1 linkage)
Methods for the linkage errors evaluation

<table>
<thead>
<tr>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Generalization in several contexts</td>
<td>- Difficulties in model definition</td>
</tr>
<tr>
<td>- “Automatic” evaluation of the errors</td>
<td>- Difficulties in verifying the model assumptions</td>
</tr>
<tr>
<td>- Simplicity</td>
<td>- Very expensive in time and resources</td>
</tr>
<tr>
<td>- Short extensibility of the results</td>
<td></td>
</tr>
</tbody>
</table>

Approaches based on mixture models

Approaches based on sampling and clerical resolution
Linkage errors: a risk in every linkage phase

In the estimation of the linkage errors one should take into account the choices done in the several steps of the linkage procedures.

Every choices of the linkage procedure may directly or indirectly introduce linkage errors. So, at each step of the procedure, precautions and tricks can be useful to reduce linkage errors.
Linkage errors: a risk in every linkage phase

*Example 1: when creating the space of the candidate pairs, the blocking method may cause missed matches*
Linkage errors: a risk in every linkage phase

Example 2: in the data pre-processing phase, if suitable tools for parsing of variables as names and addressed are not available, such variables cannot be used in the better way as matching ones and several true links cannot be detected
Bibliography


Blakely, Salmond (2102) Probabilistic record linkage and a method to calculate the positive predicted value, International Journal of Epidemiology, 2002, 31, 1246-1252


Bibliography


Heasman D. (2014) Sampling a Matching Project to Establish the Linking Quality. Survey Methodology Bulletin No. 72


Training course on record linkage and statistical matching

*Tiziana Tuoto*

*Istat*

tuoto@istat.it

*Analysis of linked data sets*
Outline

- What happens after the link status has been determined
- Statistical analyses of linked data: methods by
  - Neter, Maynes and Ramanathan, 1965
  - Scheuren and Winkler, 1995
  - Larsen and Lahiri, 2000
  - Chambers, 2009 and extensions
  - Chipperfield, Bishop and Campbell, 2011
What happens after the link status has been determined

- Record linkage often is not an objective itself, but it is an intermediate step for many applications.
- Statistical analyses based on linked data have to take into account that results can be affected by linkage errors.
- Statistical analyses based on linked data have to take into account the typical trade-off between the two kinds of errors and they have to measure their effects on the analysis results.
What happens after the link status has been determined

The usual organization of the work assigns the job of linker and data analyst to distinct people

What should the linker do to help the analyst?

What should the analyst know about the linkage and how should that information be used?
Statistical analyses of linked data

Linkage and analysis steps *should be considered part of a single statistical activity and the appropriate strategies have to be devised accordingly*.

There is an analogy between the effects of nonresponse and linkage errors (false links and missed links) on statistical analyses: often they both can invalidate standard statistical procedures.
Statistical analyses of linked data

What should the linker do to help the analyst?

For the analyst’s sake, the linker needs to provide as much linkage information as possible on the files matched so that the analyst can make informed choices in his or her work (info on matching variables, blocking procedure, matching weights or estimated linkage probabilities, clerical review steps, if it is the case)
Statistical analyses of linked data

What should the analyst know about the linkage and how should that information be used?

The analyst needs to have information like link, non link, possible link along with linkage probabilities, if available.

These info are precious also to evaluate the sensitivity of the results (e.g. regression analyses), also by adjustments and/or improved models.
**Statistical analyses of linked data**

In actual applications, the optimality of the decision rule from Fellegi and Sunter is heavily dependent on the accuracy of the estimates of the probabilities \( m \) and \( u \). Estimated parameters are (nearly) optimal if they yield decision rules that perform (nearly) as well as the same rule consisting of the actual true parameters.

The necessity to have good parameter estimates is highlighted by the way cutoff thresholds are determined, i.e. fixing error bounds.

Estimating the amount of linkage errors is not straightforward, but it is important!
Statistical analyses of linked data

Statistical analyses may be sufficiently compromised and results of standard statistical techniques could be misleading, if the resultant linked data contains a substantial proportion of false links or a significant proportion of true links are erroneously left apart (e.g. in the possible match set)
Data analysis: toy example

From Neter, Maynes and Ramanathan, 1965

A: data set with $n$ units, variables $X$ and $Z$

B: data set with $n$ units, variables $X$ and $U$

Every unit $a \in A$ is linked to one (and only one) $b \in B$

A record linkage method is applied, using as matching variables the common variables $X$
# Initial data sets

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>Z</td>
</tr>
<tr>
<td>X_{a1}</td>
<td>Z_{a1}</td>
</tr>
<tr>
<td>X_{a2}</td>
<td>Z_{a2}</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>X_{an}</td>
<td>Z_{an}</td>
</tr>
<tr>
<td>X_{b1}</td>
<td>U_{b1}</td>
</tr>
<tr>
<td>X_{b2}</td>
<td>U_{b2}</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>X_{bn}</td>
<td>U_{bn}</td>
</tr>
</tbody>
</table>
Statistical analysis on linked data: Neter et al, 1965

We are interested in analyzing the relationship between \((Z_a, U_a)\), on the \(n\) pairs \((a=1,...,n)\)

For the linkage error we actually observe \((Z_a, V_a)\) where:

\[
V_a = \begin{cases} 
u_a & \text{with probability } p \\ 
u_b & \text{with probability } q \ (a \neq b) \end{cases}
\]

and \(p+(n-1)q=1\)
The linked data set

<table>
<thead>
<tr>
<th>X</th>
<th>Z</th>
<th>V</th>
<th>U</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xa1</td>
<td>Za1</td>
<td>Va1</td>
<td>Ua1</td>
</tr>
<tr>
<td>Xa2</td>
<td>Za2</td>
<td>Va2</td>
<td>Ua2</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Xan</td>
<td>Zan</td>
<td>Van</td>
<td>Uan</td>
</tr>
</tbody>
</table>

The orange part is actually observed. The green part is the actual (not observed) U
First analysis

If we look at the statistical characteristics of the marginal variable \( V \), it can be seen that \( V \) is able to reproduce (in average) just some of the unobserved characteristics of \( U \).

\[
E(\bar{V}) = \bar{U}
\]

Sample average of \( V \) is an unbiased estimator of the sample average of \( U \)

\[
E(S_{V}^{2}) > S_{U}^{2}
\]

Sample variance of \( V \) is in average larger than the sample variance of \( U \)
How statistical relationships are influenced

Anyway, the worst effect is on the relationship parameters.

For instance, the correlation coefficient of $Z$ and $V$ is in average smoothed down with respect to the correlation coefficient of $Z$ and the actual $U$

$$| E(\rho_{Z,V}) | < | \rho_{Z,U} |$$

Why? Well, the $n$ sample observations are i.i.d.!
Comments

Reliable estimates of the probabilities of mismatch seem to be extremely important for the analyses of linked files.

Models more realistic (and more complicated) than the one by Neter et al have been defined by Scheuren and Winkler (1993, 1997) Larsen (1999), Lahiri and Larsen (2000, 2005)
Statistical analysis on linked data: Scheuren and Winkler, 1993

Scheuren and Winkler (1993, 1997) extend the approach and the model proposed by Neter et al. 1965, introducing more realistic (i.e. more complicated) assumptions

\[ V_a = \begin{cases} 
  u_a & \text{with probability } q_{aa} \\
  u_b & \text{with probability } q_{ab} \ (a \neq b)
\end{cases} \]

with \( q_{aa} + \sum_{b \neq a} q_{ab} = 1 \)
Statistical analysis on linked data: Scheuren and Winkler, 1993

Under this model, the sample average of $V$ is a biased estimator of the average of $U$. So, they propose some unbiased adjusted estimators.

The adjustment is based on two crucial assumptions:

1. the probability of being a true match is available for each pair
2. the true match is the pair with the highest matching weight (probability)
Statistical analysis on linked data: Scheuren and Winkler, 1993

Using the adjusted estimator, if the size of data increases then the MSE decreases; without adjustments, the bias rises to an unacceptable level.

The accuracy of the linkage probability estimates is crucial for the adjustment.
Statistical analysis on linked data: Larsen and Lahiri, 2005

Further extensions of the Scheuren and Winkler model, based on simulations.
Larsen (1999), Lahiri and Larsen (2000, 2005) use mixture models to improve the estimates of the matching probabilities.
Larsen (1999) finds the ML estimates and the posterior distribution estimates of the parameters of the mixture model. The difference between values is used to measure uncertainty in the relationship between records.
Statistical analysis on linked data: Larsen and Lahiri, 2005

Lahiri and Larsen (2005) propose an adjustment method for the regression model alternative with respect to the bias correction proposed by Scheuren and Winkler.

They compare the estimator variances by means of Monte Carlo simulations.
A systematic overview


In the background and assumption chapter, Chambers describes the already developed approaches.
A systematic overview

Following the approach in Scheuren et al, Chambers gives unbiased estimators of linear regression parameters under probabilistic record linkage, applying a bias correction to the Ordinary Least Squares (OLS) estimates. Their proposal is based on a ratio-type correction for the bias in the standard estimator

$$ \hat{\beta}_R = \left( \sum_q X_q^T E_q X_q \right)^{-1} \left( \sum_q X_q^T y_q^* \right) $$
A systematic overview

A different approach was defined by Larsen and Lahiri, fitting a regression relationship between the linked values $y^*$ and the values $X$, and realizing that

$$E_X(y_q^*) = E_X(A_q)E_X(y_q) = E_q X_q \beta = H_q \beta$$

The estimator becomes

$$\hat{\beta}_C = \left( \sum_q X_q^T E_q^T \Sigma_q^{-1} E_q X_q \right)^{-1} \left( \sum_q X_q^T E_q^T \Sigma_q^{-1} y_q^* \right)$$

Its optimality depends on homoscedasticity of regression errors of the linear model with modified set of explanatory variables that take into account linkage errors.
Chambers Estimators

To overcome limitation due to strong conditions of previous estimators, therefore Chambers (2009) suggests a Best Unbiased Estimator (BLUE) or its empirical (EBLUE) version.

Chambers provides also estimators:

- under the assumption of normality, using maximum likelihood
- using general estimating functions (with application to linear and logistic regression functions)
Chambers Estimators

- Furthermore, he suggests how to estimate variances of the previous estimators, including the case when linkage probabilities are estimated.

- Another extension consists in assuming that A and B do not consist of an equal number of records that link according to a 1:1 basis, but the linkage is incomplete in the sense that one of the registers in fact only covers a subset of the population fully covered by the other register. He proposes to use weighted estimating function...
Chambers Estimators: comments

Those settings are subject to strong assumptions:

- Exchangeability linkage errors model, at least into groups of records
- Equal size of linking sets (or smallest set contained in the biggest one and no interaction between sample selection process and linkage error processes)
- Linking in 1:1 constrain
- No duplicates
Chambers Estimators: further extensions


Samart (2011) extends the research to the class of linear mixed models with linked data.

Please, visit University of Wollongong Research Session Website for updates!
Analysis on linked data - loglinear models

Chipperfield, Bishop and Campbell (2011) develop a ML approach using the EM algorithm.

For instance, in contingency table, cell counts are:

\[
\tilde{n}_c = \sum_{i} \tilde{w}_{ic} = \sum_{i} \left[ w_{ic}^* \times p_{y^*} + \left( 1 - p_{y^*} \right) \tilde{\pi}_c \right]
\]

\( w_{ic}^* \) assumes value 1 if the \( i \)-th link is classified in cell \( c \).

\( p_{y^*} \) is the probability that the \( i \)-th link is correct.

\( \tilde{\pi}_c \) is the probability of the \( c \) cell.

Estimates are obtained iterating determination of \( \tilde{n}_c \) and evaluation of \( \tilde{\pi}_c \).
Analysis on linked data: Chipperfield et al (2011)

Chipperfield et al. (2011) also propose PML to include in the analyses un-matched records.

The PML is based on weighting the scores involved in estimation of the model by the inverse of the probability that a records remains unlinked, i.e. the counts are

\[ q\tilde{n}_c = \sum_i q_i \tilde{\psi}_{ic} \]

Estimation, as before, iterating estimates of \( q\tilde{n}_c \) and \( \tilde{\pi}_c \)
Chipperfield et al (2011): comments

- They illustrate the method both for the analysis of contingency table and logistic regression.
- This method overcomes the constraint that all records have to be linked and explicitly consider in the analysis the unlinked data or missed links.
- The method does not require exchangeability of linkage errors.
- It can be applied without restrictions also to cases when multiple steps are performed for linking records.
Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


