

SUPPORTING DECISION-MAKING IN THE ENTITY NORMALIZATION TASK. ON THE EXAMPLE OF PROVIDEDH.

Grzegorz Bryk ^[1], Alejandro Benito ^[2], Michał Kozak ^[1], Cezary Mazurek ^[1], Szymon Płaczaszek ^[1], Alejandro Rodríguez ^[2], Roberto Therón ^[2]
[1] Poznań Supercomputing and Networking Center, [2] University of Salamanca

Uncertainty visualization in TEI documents | The main focus of the PROVIDEDH project is geared toward the creation of visualization systems, that aid in the handling of uncertainty in Digital Humanities datasets. Our tools on the collaborative platform were developed against the background of human-centered design, with a focus on easing uncertainty annotation and visualization, promoting the use of TEI standards and making uncertainty play a more active role in the research process.

Multi-user system with support for uncertainty annotations | Our platform allows users to create research projects, upload batches of TEI files, annotate these files concurrently by multiple users, as well as identify and manage the entities appearing in the files. Users are able to create their own taxonomies of uncertainty or use the default one to make annotations. All annotations and unifications made on the platform can be rendered then to the TEI format. The users do not need to be familiar with TEI, as our platform provides them a friendly interface for annotating electronic texts. This allows a variety of users to use the platform in ways suitable to their various needs and datasets.

Automatic detection of different instances of the same entity | One of the functionalities of the platform is Entity Normalization. To achieve this, the system employs a neural network and weighted multigraphs approach to recognize entities appearing in the TEI files under different identifiers and to suggest unifications when certain similarities are detected. In the multigraphs, the nodes (representing entities) are connected by edges constructed and weighted by metadata analysis. The composition of metadata set depends on the type of entity; for a "person" it is, among others: forename, surname, gender, dates, mentions' context and mutual acquaintances — i.e. co-occurrence of unified people in the same files of the dataset.

Supporting decision-making in the unification of entities | The result of the entity normalization task can also be presented in the form of a graph, in which the nodes (representing entities and cliques of already unified entities) are connected by three types of edges. **Purple** edges show the proposed unifications between entities and cliques — the thickness of these edges illustrates the system confidence for the unification. Grey edges show co-occurrence of entities in one file, and **turquoise** edges show entities belonging to a clique. The system presents each of the proposed unifications together with the context of the occurrence of entities being unified, where the user can reject or accept it. In case of acceptance, the user must express their confidence of the unification. The system learns from users' actions and subsequent runs of the algorithm take the rejected and accepted unifications into account.

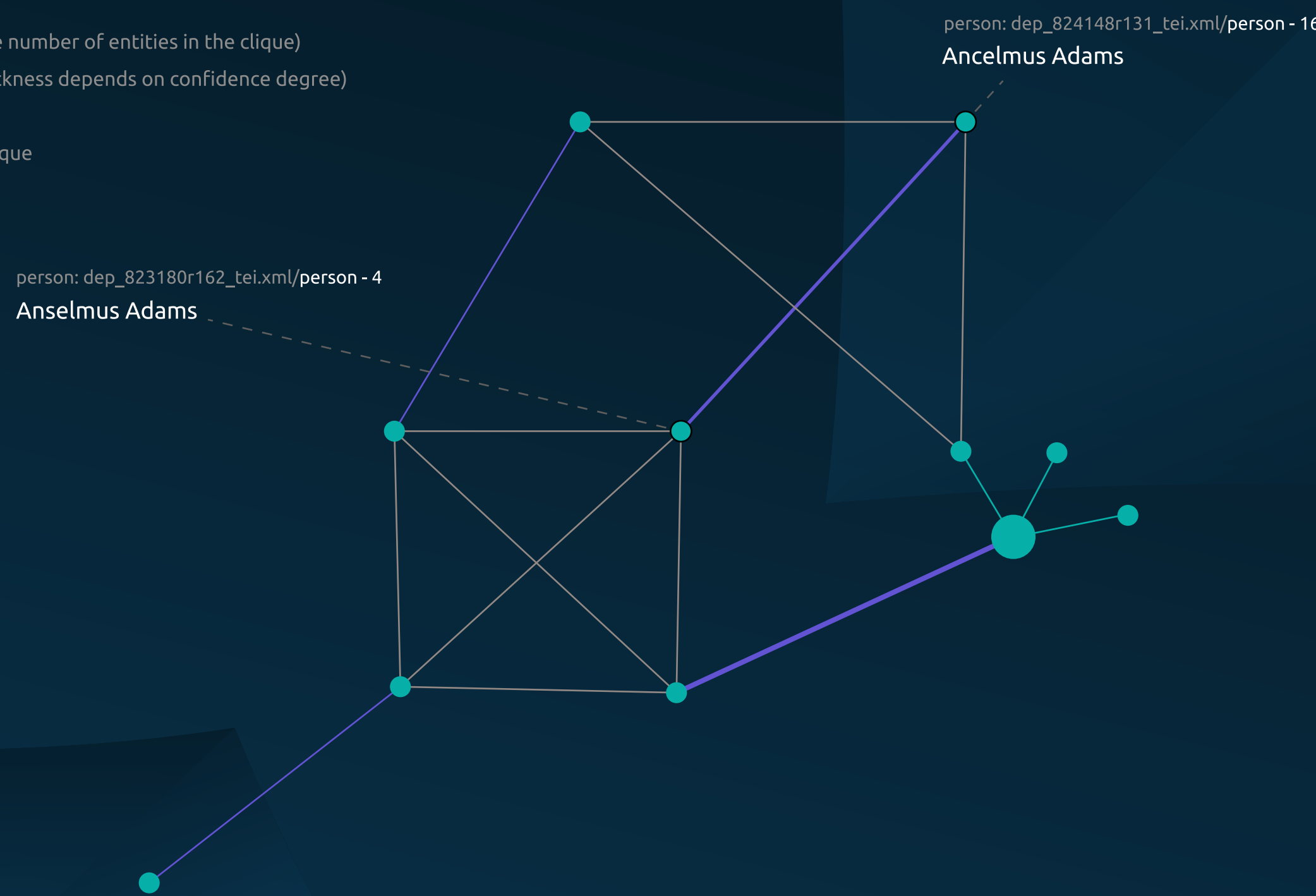
Example 1 dep_823180r162_tei.xml

[...] persons this deponents goods valued thirtie pounds were then & there burnt & taken away by them & their forces namely William lord Baron of Castle connell Patrick Purcell of Croe in the County of limericke Esquire lieutenant generall of the forces of Munster Sir Edward fitz Harris knight barronett Thomas mc Cragh of Kilbeny in the County of limerick gentleman and his son Donogh mc Cragh of the same gentleman David Roch son and heire to the lord Roch all the said parties haueing mett together with their forces (at least a thousand armed men) besides the burneing of the said Towne, then & there they murthered or caused to be murthered in a rebellious maner the wife of Anselmus Adams (& her two children) likewise William Browne & Edward Blissitt of the said Towne two english protestants were then & there eight by some of the said forces. He 327B further deposeeth that aboute Christmas last he then coming from Corke was sett vpon by some of my the lord of Muskries forces in the kinges high way and whoe forceably tooke away from this depon

person-4 (person)
Attributes
Name: Anselmus Adams
Forename: Anselmus
Surname: Adams
Occupation: Victim
Sex: 1

Legend

- Entity
- Clique (size depends on the number of entities in the clique)
- Proposed unifications (thickness depends on confidence degree)
- Entities from the same file
- Entities that belong to a clique



File name: dep_823180r162_tei.xml

dep_823180r162_tei.xml.xml / person-4

Name: Anselmus Adams
Forename: Anselmus
Surname: Adams
Occupation: Victim
Sex: 1

dep_824148r131_tei.xml / person-16

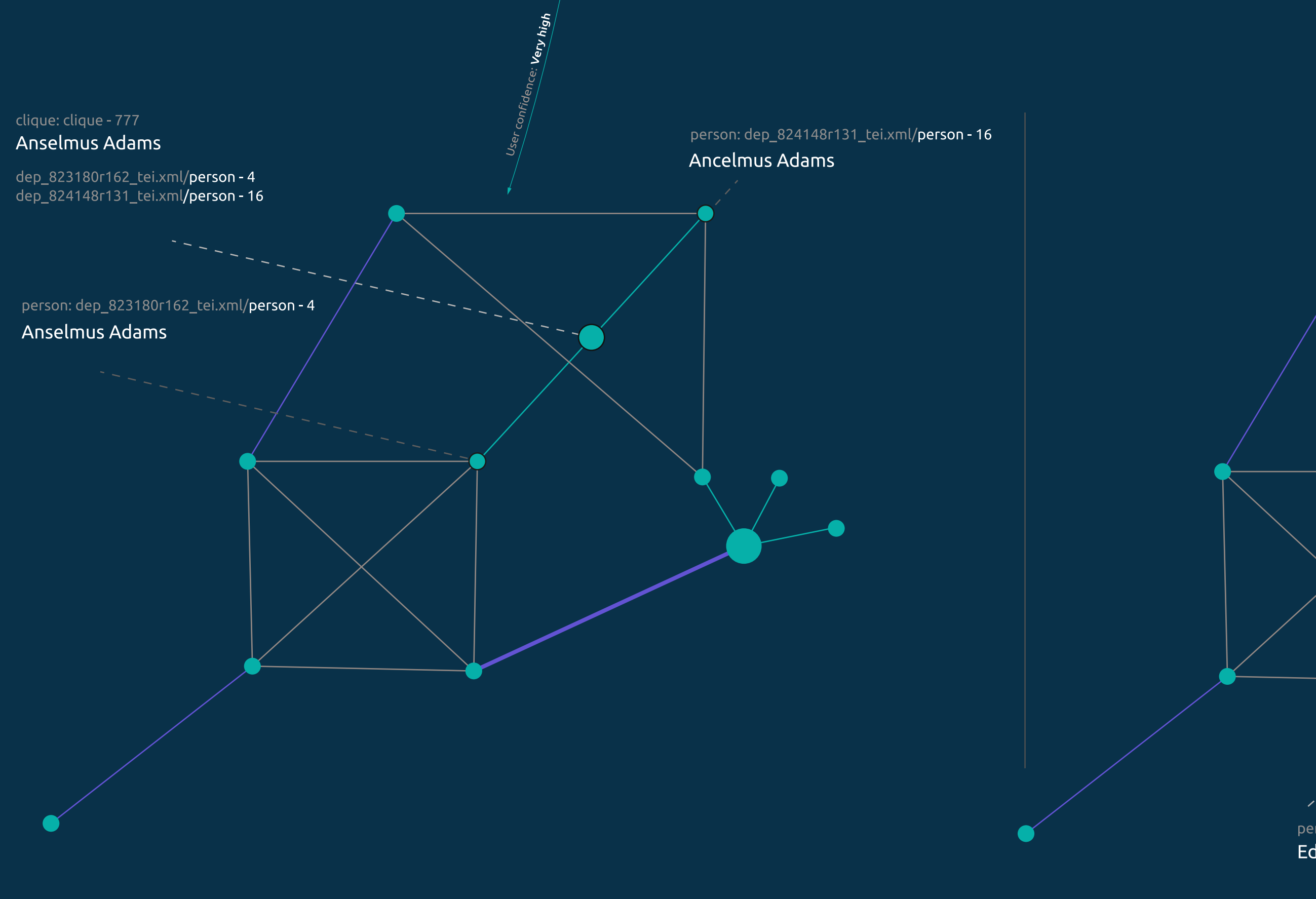
Name: Anselmus Adams
Forename: Anselmus
Surname: Adams
Occupation: Victim
Sex: 1

Confidence: 90%

Prev 2 / 4 Next

[...] mett together with their forces (at least a thousand armed men) besides the burneing of the said Towne, then & there they murthered or caused to be murthered in a rebellious maner the wife of Anselmus Adams (& her two children) likewise William Browne & Edward Blissitt of the said Towne two english protestants were then & there eight dayes before murthered by some of the said forces. He 327B further deposeeth that aboute Christmas last he then coming from Corke was sett vpon by some of my the lord of Muskries forces [...]

[...] the Right honorable the Lo: Viscont Roch all the said parties haueing mett together with there forces (at Least a thowsand armed men) came with force & armes to Michellstowne aforesaid where in a rebellious manner 1891Burnt the Towne and the Church and Likewise then & there murred the wife of Anselmus Adams (and two children of his) William Browne and Edward Blissitt and likewise then & there forceably tooke away this deponents goods Chattles and further he [...]



Example 2 dep_824148r131_tei.xml

[...] The totall of his Losses amounts to the value of fortye fiue pounds And further he sayth that his goods were taken away by the in manner followinge vizt, aboute the 14th day of ffebruary Last 164 William Lo: Barron of Castle Connell Patrick Purcell of Cro in the com of Limericke Esquire Leifttenant generall of the Forces of Munster Sir Edward fitz Harris knight & Barron Thomas McCragh of Kilbeny in Com of Limerick gentleman and his sonn & heire Donnagh mc Cragh, David Roch sonn & heire to the Right honorable the Lo: Viscont Roch all the said parties haueing mett together with there forces (at Least a thowsand armed men) came with force & armes to Michellstowne aforesaid where in a rebellious manner 1891Burnt the Towne and the Church and Likewise then & there murred the wife of Ancelmus Adams (and two children of his) William Browne and Edward Blissitt and likewise then & there forceably tooke away this deponents goods Cha

person-16 (person)
Attributes
Name: Ancelmus Adams
Forename: Ancelmus
Surname: Adams
Occupation: Victim
Sex: 1

Proposed unifications

Type	Entity	Matched entity or clique	Entities in clique
Person	dep_823180r162_tei.xml/person-3	dep_824148r131_tei.xml/person-14	
Person	dep_823180r162_tei.xml/person-4	dep_824148r131_tei.xml/person-16	
Person	dep_823180r162_tei.xml/person-6	clique-235	dep_824148r131_tei.xml/person-18 dep_821032r016_tei.xml/person-767 dep_821047r026_tei.xml/person-2865
Person	dep_823180r162_tei.xml/person-5	dep_819079r103_tei.xml/person-17	
Person	dep_811095r067_tei.xml/person-123	clique-34	dep_813352r264_tei.xml/person-244 dep_809290r178_tei.xml/person-1033 dep_815001r002_tei.xml/person-2044 dep_813244r168_tei.xml/person-344 dep_816187r115_tei.xml/person-1355
Person	dep_811095r067_tei.xml/person-99	dep_811023r004_tei.xml/person-987	

File name: dep_823180r162_tei.xml

dep_823180r162_tei.xml.xml / person-6

Name: Edward Blissitt
Forename: Edward
Surname: Blissitt
Occupation: Victim
Sex: 1

clique-235 / dep_821032r016_tei.xml/person-767

Name: Edw: Blissitt
Forename: Edw:
Surname: Blissitt
Occupation: Victim
Sex: 3

Confidence: 99%

Prev 3 / 4 Next

[...] parties haueing mett together with their forces (at least a thousand armed men) besides the burneing of the said Towne, then & there they murthered or caused to be murthered in a rebellious maner the wife of Anselmus Adams (& her two children) likewise William Browne & Edward Blissitt of the said Towne two english protestants were then & there eight dayes before murthered by some of the said forces. He 327B further deposeeth that aboute Christmas last he then coming from Corke was sett vpon by some of my the lord of Muskries forces in the kinges high way and whoe forceably tooke away from this deponent of goods and merchantable ware worth ten pounds [...]

[...] last her said husband hauing in company with him thirty men when they left the Castle of Mitchelstowne in the com of Corke & all warders of the same purposeing to bring a prey from the rebels towards the reliefe of the said Castle, but the Enemy (the forces of the Lord of Cahir) laying ambushes for them murthered & killed this deponents said husband & Edw: Blissitt Tho: Dan: Thomas Handcock John Watton George Horne Roger Laurence, Tho: Wats which parties left behinde them forty fartherlesse Children betweene them not able to releiue themselves Robert Best late of Balaporine B aforesaid this day was produced before vs as a Wittnes & being Examined & sworne vpon the holy Euangelist he deposeeth that the Contents of [...]

