

Prosjektrapport

Enhancing Digital Infrastructure for Historical Person Data (EDISH)

Tilskuddsmottaker: Stiftelsen Falstadsenteret
Samarbeidspartner: SINTEF Digital
Prosjektperiode: 13.01.2025-29.08.2025
Rapport utarbeidet av: Erlend Stav (SINTEF Digital), Shanshan Jiang (SINTEF Digital)
og Arne Langås (Stiftelsen Falstadsenteret)

Prosjektet EDISH er gjennomført av Falstadsenteret i nært samarbeid med SINTEF Digital. Det er utarbeidet oversikt over relaterte standarder og løsninger for personhistoriske databaser, definert brukerhistorier og krav, utarbeidet skisser til datamodell og presentert erfaringer fra validering med konkrete eksempler. Det langsiktige målet for prosjektet er en nasjonal forskningsinfrastruktur basert på en datamodell som håndterer utfordringer knyttet til person- og stedsnavn, koblinger mellom personer, steder, historiske hendelser og kilder, samt forsvarlig databehandling.

Rapporten inneholder følgende hoveddeler:

1. Mål og gjennomføring
2. Resultater og gevinster
3. Erfaringer og utfordringer
4. Videre arbeid
5. Vedlegg (figurer)

1. Mål og gjennomføring

Målet med spireprosjektet EDISH var å analysere eksisterende dataløsninger for historiske persondata og tilknyttede steds- og hendelsesdata, samt å utvikle en skisse til en datamodell for en nasjonal forskningsinfrastruktur.

Prosjektet er gjennomført av Falstadsenteret i tett samarbeid med SINTEF Digital. Partene har hatt flere arbeidsmøter for bl.a. utveksling av erfaringer og kunnskap, diskusjon av problemstillinger og definisjon av brukerhistorier.

I tråd med forutsetning i tildelingsbrev, ble det 26. juni 2025 holdt digitalt møte med Arkiverket (v/Kristin Jacobsen og Trine Kruse) hvor vi presenterte EDISH. Det er avholdt digitalt møte med Norsk regnesentral (Historisk befolkningsregister) om mulig samarbeid/samkjøring. Vi har også vært i dialog med Narviksenteret (Soldater.no), Norsk senter for krigsseilerhistorie v/Arkivet (Krigsseilerregisteret.no) og Karlsuniversitetet (Noráci) som har bidratt med input fra sine eksisterende databaser (oppbygging og/eller innhold).

2. Resultater og gevinster

I søknaden ble det definert fire resultater:

- R1: Oversikt over relaterte standarder og løsninger
- R2: Brukerhistorier og krav
- R3: Skisse til datamodell
- R4: Erfaringer fra validering med eksempler

Resultatene presenteres nærmere i delkapittel 2.1.–2.4. R1 tilsvarer arbeidspakke A i søknaden, mens R2, R3 og R4 samlet tilsvarer arbeidspakke B.

2.1. EDISH R1 – Oversikt over relaterte standarder og løsninger

2.1.1. Metode/fremgangsmåte

Informasjonen er innhentet fra:

1. Nøkkelordbaserte nettsøk.
2. Bidrag og fagkunnskap fra fagekspertene innen arkivfag og kulturhistorie, samt løsningsspesialister (for eksempel forskere innen ontologi- og metadatastandarder).
3. Kartlegging av eksisterende databaser som kan kobles med databasene Falstadsenteret eller assosierte partnere i EDISH forvalter, f.eks. ved integrering eller utveksling data.

2.1.2. Eksisterende databaser

Nedenfor følger en kort oversikt over databaser som Falstadsenteret forvalter, og relaterte databaser som andre forvalter.

(Forkortelser: PE = Prosjekteier EDISH. AP = Assosiert partner EDISH.)

Navn og URL	Ansvarlig etc.
Utenlandske krigsgraver i Norge https://www.krigsgraver.no/	Falstadsenteret (PE)
Norsk digitalt fangearkiv 1940-1945 https://www.fanger.no/	Arkivet (AP) Falstadsenteret (PE)
Krigsseilerregisteret https://www.krigsseilerregisteret.no/	Arkivet (AP)
Historisk befolkningsregister https://www.histreg.no/	Norsk regnesentral (AP)
Soldatregisteret https://soldater.no/	Narviksenteret (AP)
Noráci https://noraci.cz/no/hovedside/	Karlsuniversitetet, Praha (AP) NTNU (AP)
Odveleni na server https://odveleninasever.cz/	Karlsuniversitetet, Praha (AP)
Gravminner https://www.slektogdata.no/gravminner/finn-en-grav/	Slekt og Data
Våre falne https://www.vg.no/spesial/2015/vaare_falne/	VG
Snublesteiner https://www.snublestein.no/	Jødisk Museum Oslo
European Holocaust Research Infrastructure (EHRI) https://portal.ehri-project.eu/	EHRI-ERIC, Nederland
Arolsen Archives https://arolsen-archives.org/en/	International Center on Nazi Persecution, Tyskland
The Central Database of Shoah Victims' Names https://collections.yadvashem.org/en/names/	Yad Vashem, Israel

The Pan-European Deportation Database https://collections.yadvashem.org/en/deportations/	Yad Vashem, Israel
OBD Memorial https://obd-memorial.ru/html/	Russlands forsvarsdepartement, Russland

Av disse er følgende tre databaser nærmere gjennomgått i analysen i avsnitt 2.3:

- **Fanger.no**
 - **Ansvarlig:** Arkivet freds- og menneskerettighetssenter og Falstadsenteret.
 - **Lansert:** 2020.
 - **Innhold:** Norske fanger i Norge og utlandet (og på sikt over utenlandske fanger i Norge) under 2. verdenskrig.
 - **Sentrale elementer:** Personer (navn, født, død, bosted, info om fangenskap), fangesteder, historier (fagtekster), stedsregister (land, kommune, fylke), kilder og multimediafiler, lenker til skannede dokumenter på andre plattformer (bl.a. Digitalarkivet og OBD Memorial).
- **Krigsgraver.no**
 - **Ansvarlig:** Falstadsenteret.
 - **Lansert:** 2010/2018
 - **Innhold:** Utenlandske krigsgravlagte i Norge, oppdrag fra Krigsgravtjenesten i Kultur- og likestillingsdepartementet.
 - **Sentrale elementer:** Personer (navn, født, død, info om gravsted), gravplasser, kilder og multimediafiler, lenker til skannede dokumenter på andre plattformer (bl.a. Digitalarkivet og OBD Memorial).
- **Soldater.no**
 - **Ansvarlig:** Narviksenteret
 - **Lansert:** 2022
 - **Innhold:** Norske soldater under felttoget i 1940.
 - **Sentrale elementer:** Personer (navn, født, død, bosted, soldatnummer og grav), militære avdelinger, stedsregister (knyttet til født, død, bosted, krigsfangenskap).

2.1.3. Informasjons- og datamodeller

- **Felles informasjonsmodell for Person og Enhet**
 - Utgiver: Digitaliseringsdirektoratet
 - Felles datakatalog (data.norge.no)
 - <https://data.norge.no/information-models/8bfa055a-9b0c-40df-a7c2-2e58173eb2e0>
- **Felles informasjonsmodell for Adresse**
 - Utgiver: Digitaliseringsdirektoratet
 - <https://data.norge.no/information-models/c9f2b84d-6cc7-4eef-96e0-92cb887ca761>
- **NUTS / LAU – ISO standard regarding administrative divisions**
 - NUTS / LAU er en standard for administrativ inndeling som brukes i EU (inkl. kandidatland og EØS).
 - https://en.wikipedia.org/wiki/ISO_3166
 - https://en.wikipedia.org/wiki/Nomenclature_of_Territorial_Units_for_Statistics
 - <https://www.ssb.no/en/klass/klassifikasjoner/508>
- **GEDCOM: standard for slektsdata**
 - GEDCOM er et standardformat for utveksling av digitalt slektsdata, utviklet av Jesu Kristi Kirke av Siste Dagers Hellige. Standarden gjør det mulig å dele genealogisk informasjon og bruke ulike kompatible programvarer.
 - <https://gedcom.io/specifications/gedcom7-rc.pdf>

- **IDS – The Intermediate Data Structure**
 - IDS definerer et standard dataformat som brukes i en del historiske befolkningsdatabaser. I Norge er den brukt av UiT.
 - <https://hlcs.nl/article/view/9290/9796>
- **HFREG: En utvidelse av Statistisk sentralbyrås versjon av folkeregisteret (FREG)**
 - Eksakt samme format som FREG. Brukes i SSBs statistikkproduksjon, Arkivverkets saksbehandling og til utlevering til forskning fra både SSB og Arkivverket. Arkivverket har en DPIA som fastslår at dette er det eneste formatet hvor Arkivverket har lov til å koble klausulerte kilder.
 - Nettstedet histreg.no har den åpne delen av autentitetsregisteret for personer som har bodd i Norge, og brukes i et av Norges største folkeforskningsprosjekter. Her kan man følge en stor del av befolkningen fra 1800-1920 (dødsfall frem til i dag) basert på nærmere 24 mill. lenkede personposter hovedsakelig fra folketellinger, kirkebøker og dødsfall.

2.1.4. Ontologier

- **CIDOC-CRM**
 - CIDOC Conceptual Reference Model (CRM) er en kjent ontologi innen arkivdomenet. Denne blir også brukt av Arkivverket.
 - <https://cidoc-crm.org/>
- **Geonames**
 - GeoNames er en database og tjeneste med et rikt innhold av informasjon og steder.
 - De har også definert en underliggende ontologi. Geonames støtter alternative navn på steder, f.eks. for forskjellige språk og tegnsett. Det kan også spesifiseres fra og til dato for navn, og det kan merkes av om et navn er historisk eller slang. En begrensning er at det ikke er historikk for hvordan steder er knyttet til et større (administrativt) geografisk område (som fylker og land).
 - <https://www.geonames.org/>

2.1.5. Generelle standarder

- **Standarder for representasjon av tekst**
 - UNICODE er en universell standard for digital representasjon av tekst som støtter alle kjente språk, alfabeter, numeriske verdier, symboler, mm.
 - UTF8 er den mest brukte representasjonen av tegn i UNICODE, og bruker 1-4 byte per tegn. Den er bakoverkompatibel med ASCII.
 - <https://home.unicode.org/>
- **Standarder for annotering av språk og script**
 - IETF BCP 47. IETF står for Internet Engineering Task Force, og BCP for Best Current Practice). BCP 47 er sammensatt av andre standarder, inkludert ISO 639 for angivelse av språk og ISO 15924 for script.
 - <https://www.rfc-editor.org/info/bcp47>
 - https://en.wikipedia.org/wiki/IETF_language_tag

2.1.6. Andre aktører og mulige samarbeidspartnere

Det er bygd nettverk med internasjonale aktører som kan være aktuelle for videre prosjektsamarbeid ved å delta på følgende:

- **"Nazi Persecution – Person Data and Data Standards".**
 - I oktober 2024, før formell prosjektstart, deltok Falstadsenteret og Arkivet på konferanse i München/Dachau, arrangert av Leibniz Institute for Contemporary History og Dachau Concentration Camp Memorial Site.
 - <https://www.hsozkult.de/event/id/event-143657?language=en>

- **EHRI-workshop i Riga**

- I mai 2025 deltok Falstadsenteret på workshop i Riga for potensielle nye deltakere i EHRI-ERIC fra Skandinavia, Baltikum og Ukraina. EHRI (European Holocaust Research Infrastructure) gikk i 2025 over til å bli en permanent organisasjon i form av en ERIC (European Research Infrastructure Consortium).

På begge samlingene møtte vi stor interesse for det arbeidet som er lagt ned gjennom våre eksisterende databaser og for initiativet med EDISH.

2.2. EDISH R2 – Brukerhistorier og krav

Dette kapitlet beskriver de viktigste brukerhistoriene og kravene som er identifisert og definert i prosjektet, med særlig vekt på utfordringer knyttet til representasjonen av personer og deres livshendelser (personnavn, steder og hendelser).

Brukerhistorier fungerer som en strukturert metode for å fange opp brukerperspektiver og oversette dem til presise, kontekstsensitive krav. En brukerhistorie er en kort, brukersentrert beskrivelse av en funksjon eller et behov, uttrykt fra sluttbrukerens perspektiv. En brukerhistorie uttrykkes vanligvis i følgende format:

Som en <type bruker> ønsker jeg <mål> slik at <nytte/resultat/årsak>

I denne omgang er det identifisert brukerhistorier for følgende primære brukertyper ettersom de er mest relevante for våre prosjektmål (fokus)/for testing etterpå:

- Sluttbruker (publikum, inkl. forsker som bruker dataene)
- Dataforvalter (de som eier og vedlikeholder informasjonen i databaser/systemer, kan også være forsker, inkl. historiker som registrerer data)

I tillegg har vi identifisert følgende sekundære brukertyper som det kan være aktuelt å identifisere brukerhistorier for ved en senere anledning.

- Standardiseringsorganer og andre prosjekter (for bruk av datamodell)
- Systemutviklere som skal gjenbruke eller kombinere data fra ulike databaser

Beskrivelsen inkluderer i tillegg kommentarer, herunder henvisninger til fire relevante eksempler som viser kompleksiteten i de historiske dataene: Ivan Podatsjev, Milla Klein, Jan Hammer og Milorad Perić. Eksempelene er beskrevet nærmere i avsnitt 2.2.4, og to av dem er tatt med inn i validering i avsnitt 2.4.3. Tabellene har også et felt for å sette prioritet – dette er for å kunne sette prioritet ved videre arbeid og feltet er ikke benyttet i denne omgang. Brukerhistoriene er identifisert i samarbeid mellom forskere fra SINTEF og domeneeksperter fra Falstadsenteret.

Deretter presenterer vi noen krav som supplerer brukerhistoriene. Noen krav kan være knyttet til datamodeller, andre til selve systemet (databasen). Vi har valgt å fokusere på krav til datamodeller som kan understøtte de identifiserte behovene.

2.2.1. Brukerhistorier for sluttbruker

ID	Beskrivelse	Kommentarer	Pri
SLB-01	Som sluttbruker ønsker jeg å søke etter en person basert på navn på tvers av systemer og landegrenser slik at jeg kan få en oversikt over all historisk informasjon relatert til personen	Navnesøk etter slektninger, kjente personer, folk fra eget lokalmiljø etc.	

SLB-02	Som sluttbruker ønsker jeg å hente ut alle livshendelser for en person slik at jeg kan se de på en tidslinje og over geografiske lokasjoner		
SLB-03	Som sluttbruker ønsker jeg at systemet skal kunne finne informasjon uavhengig av opprinnelig språk, alfabet og stavemåte brukt i kilden		
SLB-04	Som sluttbruker ønsker jeg å se relasjoner mellom personer	Familiære relasjoner, andre relasjoner.	
SLB-05	Som sluttbruker ønsker jeg å se nøkkelinformasjon om personer som gir delvis treff på mitt søk slik at jeg kan finne ut om en av de er personen jeg ønsker å finne.		
SLB-06	Som sluttbruker ønsker jeg å kunne foreslå endringer eller tillegg til informasjon jeg finner slik at informasjonen blir mest mulig korrekt og dekkende.		
SLB-07	Som sluttbruker ønsker jeg å kunne søke på geografiske steder (f.eks. via kart) og få treff i personer og andre relaterte entiteter slik at jeg kan få oversikt over historisk informasjon relatert til stedet.		
SLB-08	Som sluttbruker ønsker jeg å bla gjennom, filtrere og søke på tema slik at jeg kan finne all informasjon som relaterer seg til temaet.		
SLB-09	Som sluttbruker ønsker jeg å se hvilke kilder som ligger til grunn for den registrerte informasjonen, få tilgang til kildene, og få støtte til å vurdere dem.		

2.2.2. Brukerhistorier for dataforvalter

ID	Beskrivelse	Kommentarer	Pri
DF-01	Som dataforvalter ønsker jeg å kunne koble kilder til riktig person slik at vi får et rikere historisk bilde av personen		
DF-02	Som dataforvalter ønsker jeg å kunne registrere kilder i et hierarki og laste opp kildemateriale eller lenke til tredjeparts kilder, slik at vi holder oversikt over hvilke kilder som finnes og som er gjennomgått.		
DF-03	Som dataforvalter ønsker jeg å kunne registrere hvordan kilden er tolket og eventuell usikkerhet slik at det er mulig å danne seg et bilde av hvor pålitelig informasjonen er.	Er tolking og registrering gjort av fagfolk, frivillige, AI.	
DF-04	Som dataforvalter ønsker jeg å få støtte av systemet til å identifisere og markere om informasjon skal være åpen, slik at vi tar vare på personvern og etikkhensyn både for registrerte	Kan gjelde «kategorisering» av personer, helseopplysninger.	

	personer og deres slektninger, og til de som bidrar med informasjonen gjennom registrering eller innsending.		
DF-05	Som dataforvalter ønsker jeg å kunne registrere navnehistorikk (med tidspunkt for endringer), aliaser og alternative skrivemåter for navn med metainformasjon slik at personen er mer rikt beskrevet og slik at det er lettere å finne personen ved søk.	Se eksempler Ivan og Milla.	
DF-06	Som dataforvalter ønsker jeg at steder er beskrevet med koordinater og historikk for stedsnavn og administrativ tilhørighet slik at stedsangivelse kan være uavhengig av historiske endringer.	Se eksempel Jan.	
DF-07	Som dataforvalter ønsker jeg å kunne registrere livshendelser for en person, koble disse sammen og knytte de til steder og tema.		
DF-08	Som dataforvalter ønsker jeg å kunne tidfeste livshendelser til et konkret tidspunkt, et sett med alternativer eller til en periode.	Se eksempel Milla.	
DF-09	Som dataforvalter ønsker jeg å registrere bygninger / installasjoner / fartøy / etc og knytte disse til steder. Noen av disse kan eksistere bare i et visst tidsrom eller kan forflytte seg geografisk.	Se eksempel Milorad.	
DF-10	Som dataforvalter ønsker jeg å utveksle data med andre aktører (f.eks. en dataforvalter i et annet land) med filtrering på alt som gjelder en person, en lokasjon, etc.		

2.2.3. Krav

ID	Beskrivelse	Kommentarer	Bruker-historier	Pri
FK-01	Datamodellen må kunne representere koblinger mellom livshendelser og kilder.		DF-01	
FK-02	Koblinger må kunne inneholde informasjon om hvordan kilden er tolket og hvor pålitelig dette er.		DF-02	

2.2.4. Eksempler

Nedenfor presenteres eksempler på personposter som fins i flere av basene, og som inneholder komplekse opplysninger. Av disse er eksemplene Ivan Podatsjev og Milla Klein omtalt ytterligere i avsnitt 2.4.3 (validering).

Eksempel 1: Ivan Gerasimovitsj Podatsjev

Ulike språk/alfabet for representasjon av samme navn. Person oppført i ulike databaser.

- **Navn:** Ivan Gerasimovitsj Podatsjev (norsk transkripsjon), Iwan Gerasimowitsch Podatschew (tysk transkripsjon), Иван Герасимович Подачев (russisk). Samme navn, skrevet på ulike måter. I tillegg kommer direkte transkripsjon av navn i skriftlige kilder.
- **Hendelsesforløp:** Tatt til fange i krig, sendt til Norge som krigsfange, død i fangenskap. Gravlegging, flytting av graver.
- **Oppføringer:** [Krigsgraver](#). Skal på sikt også registreres på Fanger.
- **Oppføringer i andre baser:** OBD Memorial (fangekort).
- Se også eget punkt under validering i 2.4.3.

Eksempel 2: Milla Klein

Ulike navn til ulike tider. Stedsnavn som har skiftet administrativ tilhørighet og som har ulike navn på ulike språk. Person oppført i ulike databaser, og som er relatert til andre personer i samme base. Relaterte personers oppføringer i andre databaser.

- **Navn:** Milla Glück (fødenavn), Milla Klein (navn som gift), Malka Gottfried (hebraisk fornavn, etter flytting til Israel).
- **Steder:** Født i det som i dag er byen Saldus i Latvia, tidligere kjent som Frauenburg på tysk.
- **Hendelsesforløp:** Internering, fangenskap, flukt til Sverige.
- **Oppføringer:** [Fanger](#), [Histreg](#).
- **Personlenker:** Hennes ektemann [Wulf Klein](#) og deres to barn er også oppført i Fanger.no.
- **Oppføringer i andre baser:** Ingen oppføringer for henne, men for andre familiemedlemmer på Yad Vashem, Snublesteiner og Våre falne.
- Se også eget punkt under validering i 2.4.3.

Eksempel 3: Jan Hammer

Ulike varianter av navnet på ulike språk. Stedsnavn som har skiftet administrativ tilhørighet og som har ulike navn på ulike språk. Person oppført i ulike databaser.

- **Navn:** Jan Hammer (tsjekkisk navn), Johan/Johann Hammer (navnet han er registrert med i tyskspråklige kilder).
- **Steder:** Født i Štarnov som lå i Tsjekkoslovakia (etablert 1918) da han ble født i 1922. I tyskspråklige kilder brukes skrivemåten Starnau. Ligger i det området som fra 1939 var *Protektoratet Böhmen-Mähren*. Etter 2. verdenskrig igjen Tsjekkoslovakia, fra 1993 Tsjekkia.
- **Hendelsesforløp:** Arbeider i Organisation Todt. Fluktforsøk, tatt til fange, satt i fangenskap en periode 1942-43.
- **Oppføringer:** [Fanger](#), [Histreg](#), [Noráci](#)

Eksempel 4: Milorad Perić

Ulike språk/alfabet for representasjon av navnet. Stedsnavn som har skiftet administrativ tilhørighet. Person oppført i ulike databaser.

- **Navn:** Milorad Perić; etternavnet er skrevet Peritsch i tyskspråklige kilder. Født i dagens Serbia hvor det både brukes latinsk og kyrillisk alfabet; med kyrilliske bokstaver skrives navnet Милорад Перић.
- **Steder:** Født i Vojilovo, Golubac i dag i Serbia; da han ble født i 1923 var det kongeriket Jugoslavia.
- **Hendelsesforløp:** Tatt til fange, sendt til Norge (skipstransport med mange andre), fangenskap på Falstad, død i fangeleiren og gravlagt i Falstadskogen. Nåværende gravsted ukjent.
- **Oppføringer:** [Fanger](#), [Krigsgraver](#).

2.3. EDISH R3 – Skisse til datamodell

2.3.1. Metode

I arbeidet med å skissere datamodellen har vi først studert datamodellene til tre av de eksisterende databasene, Fanger (fanger.no), Krigsgraver (krigsgraver.no) og Soldater (soldater.no), og sett på hvordan disse støtter opp om brukerhistoriene definert i avsnitt 2.2 (EDISH R2). Disse tre databasene er presentert i 2.1.2. Vi har hatt tilgang til fulle databaseskjema for Fanger og Krigsgraver med tilhørende beskrivelse, og skjema for deler av databasen til Soldater. Vi har også undersøkt Historisk befolkningsregister, Krigsseilerregisteret og Noráci (tsjekkiske tvangsarbeidere i Norge), men har av kapasitetshensyn begrenset analysen til de tre førstnevnte.

Som en avgrensning, har vi fokusert på hvordan databasene representerer navn på personer og steder, variasjoner av navn som inkluderer feilstaving, alias, navn på forskjellige språk og i ulike alfabet, og navneendringer over tid. Vi har også sett på representasjon av hendelser for personer og tidfesting av disse.

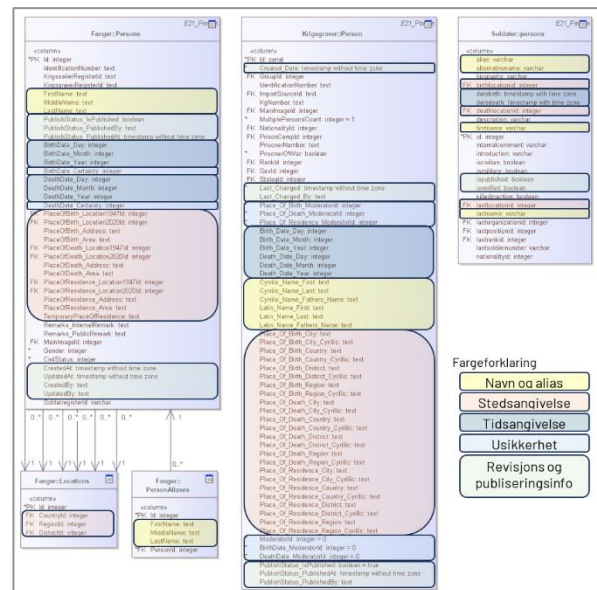
Vi har videre studert noen av standardene og løsningene identifisert i avsnitt 2.1. (EDISH R1) og hvordan disse adresserer de aktuelle problemstillingene, og vist et eksempel på hvordan felter fra en av databasene kan annoteres med en relevant ontologi.

På bakgrunn av dette har vi så skissert en konseptuell modell som dekker noen av de sentrale problemstillingene, se avsnitt 2.3.4.

2.3.2. Datamodeller i de eksisterende databasene

I dette avsnittet ser vi på hvordan databasene Fanger, Krigsgraver og Soldater håndterer personnavn, steder/stedsnavn, tidsangivelser, usikkerhet, og revisjons-/publiseringsinfo og språk/skript.

Figuren til høyre (figur 1) viser et utvalg av sentrale tabeller fra databasene, og fargede rammer er brukt for å fremheve informasjon som angir personnavn, steder/stedsnavn, tid og revisjon/publisering.



Figur 1. Se større versjon i vedlegg.

Personnavn:

- Fanger:** Navn er håndtert med at «primærnavn» har felter for fornavn, mellomnavn (=farsnavn eller navn av etternavnstype) og etternavn direkte i persontabellen. I tillegg kan et vilkårlig antall alias med samme felter legges til i tabellen PersonAliases.
- Krigsgraver:** To sett med felter i persontabellen for fornavn, etternavn og fars etternavn – et i latinsk alfabet og et i kyrillisk alfabet. Det siste brukes utelukkende for falne fra tidligere Sovjetunionen. Det er ikke noen tabell eller felter for registrering av aliaser.
- Soldater:** Navn registreres med fornavn og etternavn. I tillegg er det et felt kalt alias og et felt kalt «alternativename» i persontabellen.

Steder og stedsnavn:

- **Fanger:** Har stedsangivelser for fødsel, død og bosted som er håndtert ved at adresse og område kan angis som fritekst i persontabellen, og i tillegg kobles adressen til en post i Locations tabellen som angir land, region og distrikt (i form av videre referanser til tabeller for hver av disse). I tillegg er det et fritekstfelt for midlertidig adresse (uten kobling til Locations). Stedsangivelsene kan knyttes både til fylkes- og kommuneinndeling pr. 1947 og 2020.
- **Krigsgraver:** Har tilsvarende stedsangivelser for tilsvarende som Fanger, men bruker felter med fritekst for by, region, distrikt og land. Feltene er duplisert med varianter for latinsk og kyrillisk alfabet.
- **Soldater:** Har stedsangivelse for fødsel, død, og siste lokasjon. Nærmere beskrivelse av struktur for lokasjon var ikke tilgjengelig i denne omgang.

Tidsangivelse:

- **Fanger:** Tidsangivelse for fødsel og død er angitt i persontabellen, med felter for dag, måned og år, samt et felt for å angi usikkerhet.
- **Krigsgraver:** Tilsvarende løsning som Fanger.
- **Soldater:** Dato for fødsel og død er angitt med felter av type timestamp i persontabellen.

Usikkerhet:

- **Fanger:** Oppdelingen av dato i felter for dag, måned og år er gjort slik for å kunne la felter være blanke i tilfelle man ikke kjenner eksakt dato. Tilknyttet tidsangivelsen av fødsel og død er det også et felt som angir hvor sikker angivelsen er.
- **Krigsgraver:** Tilsvarende løsning som Fanger. I tillegg kan man også for stedsangivelsene oppgi usikkerhet knyttet til disse.
- **Soldater:** Ingen direkte felter for dette er identifisert.

Revisjons- og publiseringsinfo:

- **Fanger:** Inneholder informasjon om når en person er opprettet, sist endret og av hvem, og om posten er publisert, og i så fall når og av hvem.
- **Krigsgraver:** Tilsvarende løsning som Fanger med unntak av at det ikke er et felt for hvem som opprettet posten.
- **Soldater:** Har felter for om posten er publisert og for om den er verifisert.

Språk og script:

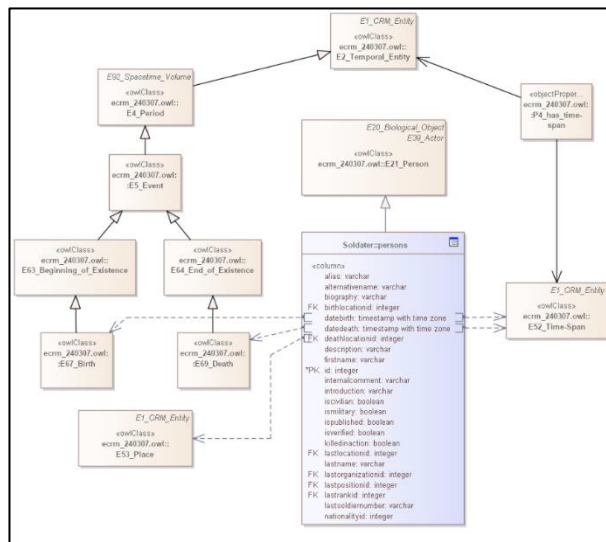
- **Fanger:** Ingen identifisert støtte for ulike språk og script.
- **Krigsgraver:** Har duplisert alle felter som angir personnavn og steder med varianter for latinsk og kyrillisk alfabet.
- **Soldater:** Ingen identifisert støtte for ulike språk og script.

2.3.3. Mapping til ontologi

En mulig fremgangsmåte for å oppnå interoperabilitet mellom datamodeller er å gjøre en mapping av hver av datamodellene til en etablert ontologi.

Innen arkivdomenet er en av de kjente ontologiene CIDOC Conceptual Reference Model (CRM), som også omtales i EDISH R1. Som et eksempel viser vi i figuren til høyre (figur 2) hvordan fødsel og dødsdato, samt sted for død fra datamodellen i soldater.no kan mappes konsepter fra CIDOC: hendelsen «Death» og dets «Timespan» som det arver fra «Temporal_Entity».

En full mapping direkte fra datamodellene til en ontologi er imidlertid et omfattende arbeid, og vi har i denne omgang sett det som mer hensiktsmessig å starte med å definere en konseptuell modell.



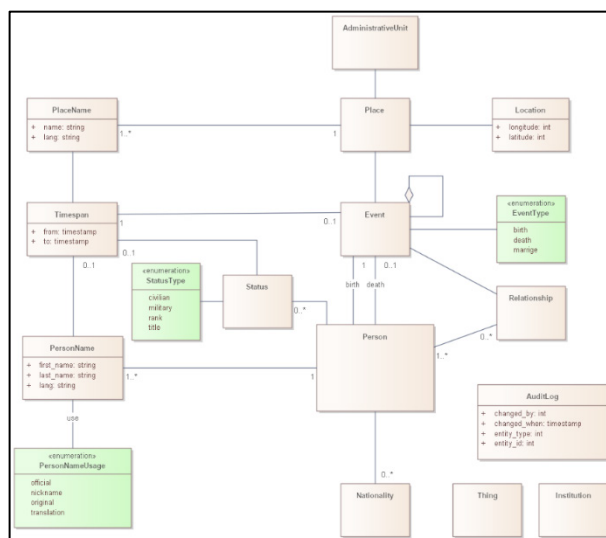
Figur 2. Se større versjon i vedlegg.

2.3.4. Skisse til konseptuell modell

Figuren til høyre (figur 3) viser en skisse til konseptuell modell som håndterer noen av de sentrale problemstillingene i EDISH, inklusive håndtering av person og stedsnavn med historikk over tid. Navn kan være bygd opp og sortert på ulike måter i forskjellige kulturer, og videre arbeid må se nærmere på hvordan dette kan håndteres. Merk at alle «Enumerations» er å betrakte som eksempler og ikke som fullstendige lister og at feltyper kun er indikasjoner på type datainnhold hvor nærmere vurderinger må gjøres ved valg av datatyper i en realisering.

Språk og script

Tekst, inkludert person- og stedsnavn bør representeres på en måte som håndterer ulike språk og tegnsett. Dette kan håndteres ved hjelp av Unicode, hvor UTF8 format er anvendt.



Figur 3. Se større versjon i vedlegg.

I tillegg til selve teksten vil det være nyttig å ta vare på metainformasjon om hvilket språk og script som er brukt i teksten. Dette kan gjøres ved bruk av standarden IETF BCP 47. IETF står her for Internet Engineering Task Force, og BCP for Best Current Practice). BCP 47 er sammensatt av andre standarder, inkludert ISO 639 for angivelse av språk og ISO 15924 for script. Som vist i PlaceName og PersonName i figuren er det definert et felt «lang» for å spesifisere dette – det er samme fremgangsmåte som brukes for web-sider.

Kjente mangler ved modellen (utgangspunkt for videre arbeid)

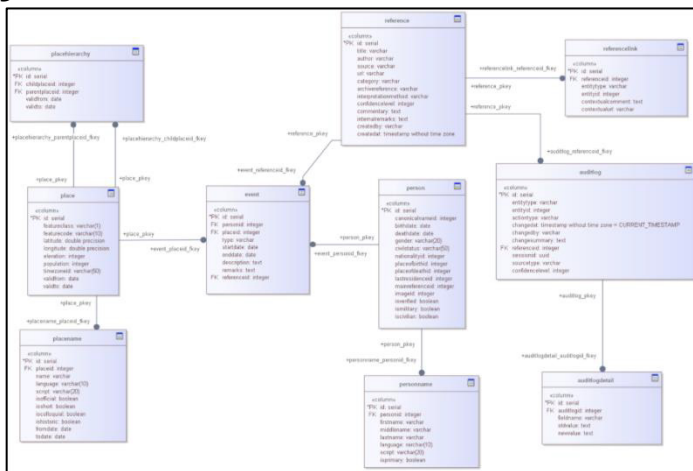
- **Forenkling:** Modellen er med hensikt forenklet til å fokusere på noen få problemstillinger.
- **Kilder:** Representasjon av kilder og referanser til disse er ikke håndtert.
- **Usikkerhet:** Modellen dekker foreløpig ikke spesifisering av usikkerhet for informasjonen. Usikkerhet kan bl.a. være knyttet til tidsangivelser og steder slik som det delvis er støttet i de nåværende databasene, men også til tolkning og pålitelighet av kilder.
- **Utvidelser:** Det mangler håndtering av historikk for noen av elementene, inklusive AdministrativeUnit. Institusjoner, bygninger og ting er ikke detaljert i modellen, og det er mulig modellen burde introdusere en felles abstraksjon som objekt eller tilsvarende som

person, sted, administrativ enhet, bygninger og ting arve fra, og hvor historikk knyttes til objekt.

2.3.5. Eksempel: Skisse til datamodell generert av ChatGPT

Som et eksperiment har vi også gitt en tekstlig beskrivelse av problemstilling til ChatGPT og bedt om å få generert en datamodell, se figur til høyre (figur 4).

Formålet med eksperimentet var å se om hvor godt den foreslåtte datamodellen generert av ChatGPT ville samsvare med den konseptuelle modellen. Et godt samsvar kan indikere at den konseptuelle modellen er i samsvar med beste praksis, mens ulikheter kan indikere noe man bør se nærmere på f.eks. ved at noe er uteglemt eller at de er en alternativ representasjonsmåte som kan vurderes.



Figur 4. Se større versjon i vedlegg.

Den genererte modellen samsvarer med den konseptuelle modellen ved at sentrale konsepter som Person, Place, Event, PersonName, PlaceName og AuditLog er til stede i begge. Den genererte modellen har et element for PlaceHierarchy mens vi i den konseptuelle modellen har et mindre detaljert konsept for AdministrativeUnit. Den genererte modellen har ikke tatt med noe eget konsept for Timespan. Den har i stedet lagt til felter som representerer dette i entitetene knyttet til Place, PlaceName, PlaceHierarchy og Event, men ikke i PersonName. Ettersom den genererte modellen er en mer full datamodell så inneholder den også langt flere felter, men det er ikke ennå hensiktsmessig å gå gjennom disse i detalj ettersom den konseptuelle modellen ikke er detaljer på det samme nivået. Overordnet kan vi konkludere med at det genererte forslaget fra ChatGPT samsvarer godt med den konseptuelle modellen på et overordnet nivå, og at vi bør se litt nærmere på Timespan-konseptet og håndtering av hierarki av steder og administrative enheter i den konseptuelle modellen.

Merk at den genererte modellen kun er egnet til inspirasjon og mulig eksperimentering, at den ikke er validert utover hva som er beskrevet i dette avsnittet, og at den ikke på noen måte må betraktes som en «fasit» for implementasjon.

2.4. Edish R4 - Erfaring fra validering med eksempler

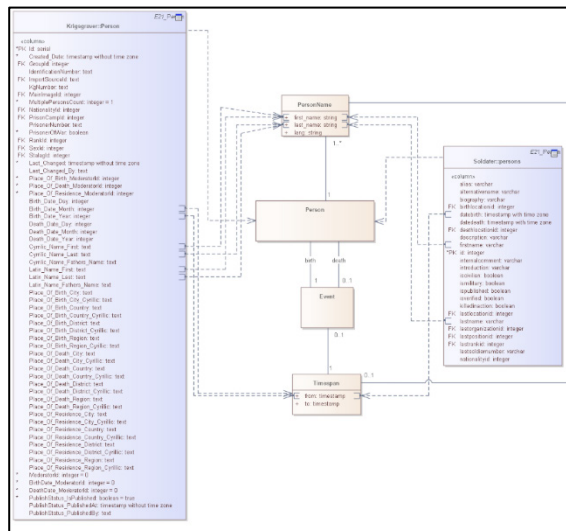
2.4.1. Metode og fremgangsmåte

I dette arbeidet har vi tatt utgangspunkt i den konseptuelle modellen fra EDISH R3 og brukerhistoriene definert i R2. En delvis validering av den konseptuelle modellen er gjort med to fremgangsmåter:

- Først har vi gjort en mapping fra utvalgte felter i to av de opprinnelige databasene til konsepter og felter fra den konseptuelle databasen.
- Deretter har vi illustrert hvordan den konseptuelle modellen kan brukes ved å illustrere instanser av konseptene med informasjon fra noen aktuelle eksempler fra databasene fylt inn. Vi refererer også til hvilke brukerhistorier fra R2 disse eksemplene helt eller delvis dekker.

2.4.2. Mapping mellom konseptuell modell og datamodeller fra eksisterende databaser

Figuren til høyre (figur 5) viser hvordan noen av de sentrale feltene fra tabellene i krigsgraver og soldater som tilsvarer person kan mappes over til elementer i den foreslåtte konseptmodellen. Mappingen er begrenset til noen av de mest sentrale elementene, men illustrerer hvordan utvalgte felter fra person-tabellen i krigsgraver og soldater kan representeres med konseptene Person, PersonName, Event og Timespan.



2.4.3. Validering med instanser fra den konseptuelle modellen

Figurene under illustrerer hvordan den konseptuelle modellen er tenkt brukt til å beskrive noen eksempler fra de opprinnelige databasene. Vi kobler også eksemplene til brukerhistoriene fra EDISH R1 som de helt eller delvis illustrerer. I figuren er navn på konseptene angitt i fane, og feltnavn er angitt i øverste rad av kolonnen i tabellene eller med <feltnavn>: før feltverdi. Verdier i "" er fri annotering for info uten felt i modellen.

Figur 5. Se større versjon i vedlegg.

Eksempel 1: Ivan Podatsjev

Den første figuren viser et eksempel med informasjon fra databasen Krigsgraver for Ivan Podasjev med representasjon av flere navn på ulike språk og skrevet i ulike alfabet. En utfordring som ikke støttes av konseptmodellen nå er at begge navnene som er skrevet i latinsk alfabet også er offisielle navn i språkene de er oversatt til.

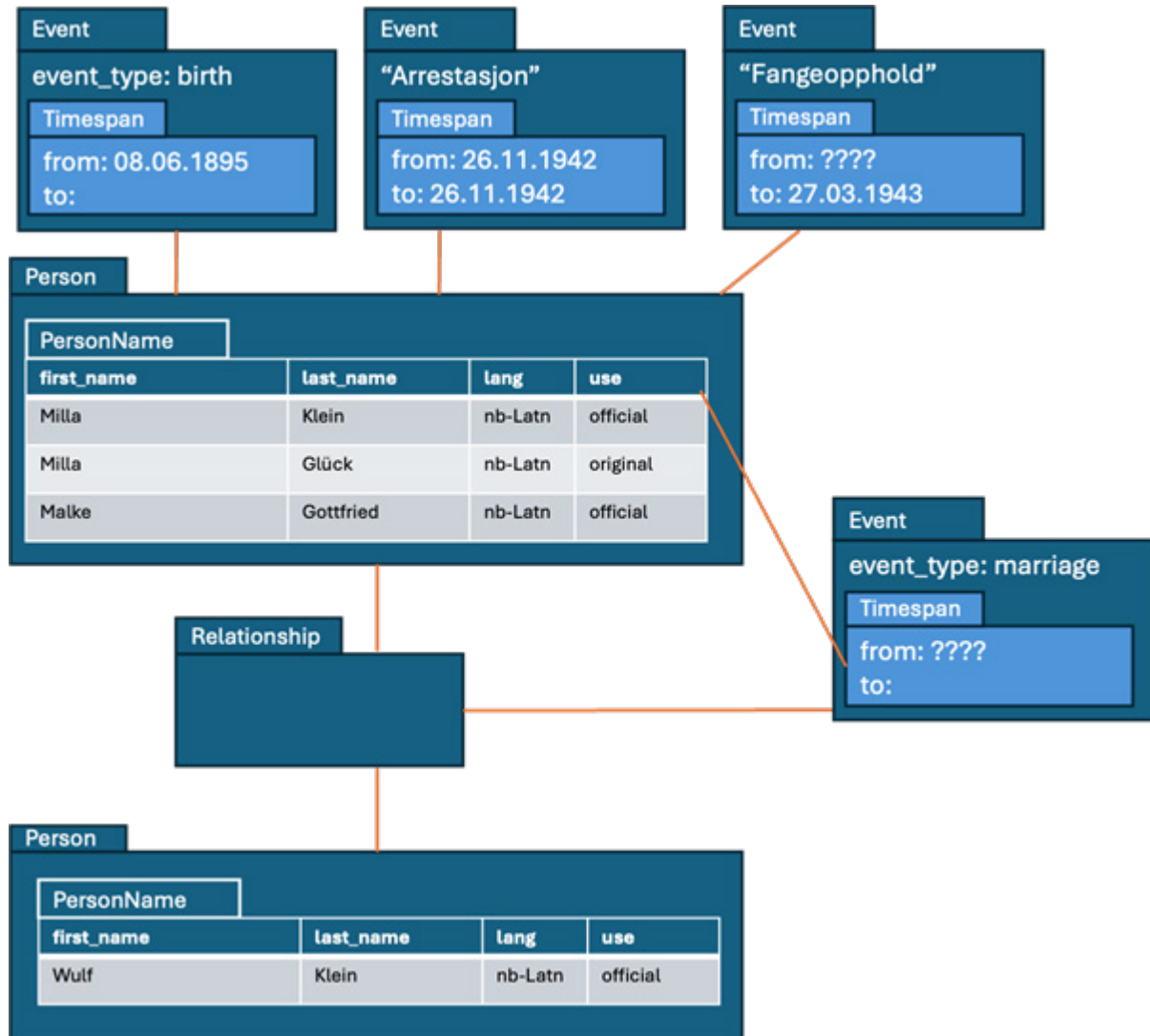
Brukerhistorier: SLB-02, SLB-03, SLB-04, DF-05, DF-07, DF-08

Person			
PersonName			
first_name	last_name	lang	use
Иван Герасимович	Подачев	ru-Cyrl	official
Ivan Gerasimovitsj	Podatsjev	nb-Latn	translation
Iwan Gerasimowitsch	Podaschew	de-Latn	translation

Eksempel 2: Milla Klein

Den andre figuren viser et eksempel med informasjon fra databasen Fanger for Milla Klein (født Glück). Figuren illustrerer hendelser i hennes liv (fødsel, arrestasjon, fangeopphold og giftemål). Det siste er også knyttet til opprettelsen av en relasjon mellom henne og hennes mann, Wulf Klein, og til en av hennes navneendringer som følge av giftemål.

Brukerhistorier: SLB-03, DF-05.



3. Erfaringer og utfordringer

En erfaring fra EDISH er at det er nyttig å starte arbeidet fra en konseptuell modell. Modellen gir en oversiktlig fremstilling som gir mulighet til å diskutere og prøve ut konseptene ved mapping og datamodeller, og ved instansiering for å studere konkrete eksempler. Mapping til ontologi kan være en omfattende og tidkrevende jobb, men det vil være mer overkommelig og oversiktlig å gjøre en slik mapping fra en konseptuell modell til en ontologi enn å gjøre tilsvarende fra en mer detaljert datamodell til en ontologi.

Her er en liste over noen konkrete problemstillinger og mangler ved konseptmodellen som har kommet frem gjennom eksemplene på instanser i forrige underkapittel:

- **Datoer:** Det bør enten legges til et konsept for å tidfeste hendelser til en enkelt dato, ellers bør konseptet Timespan (fra-til) justeres for å bedre støtte f.eks. hendelser som født, arrestert, gift og død. Dette bør også støtte usikre/upresise/motstridende datoer.
- **Event:** Det bør sannsynligvis legges til felter for navn og/eller tekstlig beskrivelse til hendelser. Event_type kan kanskje også utvides med flere verdier eller gjøres dynamisk. Dette illustreres i eksemplene i figurene hvor arrestasjon og fangeopphold er skrevet inn uten at de har et tilsvarende felt i konseptmodellen. Konseptmodellen har ikke skissert hvordan ulike hendelser kan henge sammen i komplekse hendelsesforløp.
- **Giftermål:** Hører datoene til relasjon og/eller event? Hva med ekteskapets "avslutningsdato" (Wulfs død)?
- **Navn:** Relasjonen use mellom PersonName og PersonNameUsage bør utvides til å tillate flere verdier per navn, jf. kapittel 2.3.4 og 2.4.3. I eksemplet Ivan Podatsjev er det snakk om tre ulike måter å framstille samme navn på; i eksemplet Milla Klein er det tre ulike navn (knyttet til ulike perioder i livet).

Eksemplene i avsnitt 2.4 viser at instanser av den konseptuelle modellen er egnet til å dekke noen av de identifiserte brukerhistoriene helt eller delvis. Innenfor rammene av EDISH var det ikke rom for å få til en full dekning av brukerhistoriene eller en ny iterasjon med oppdatering av konseptmodellen, men dette er noe vi ønsker å gjøre i videre arbeid.

4. Videre arbeid

Forprosjektet har synliggjort behovet for en nasjonal og europeisk forskningsinfrastruktur for historiske persondata, og Falstadsenteret vil i samarbeid med SINTEF Digital videreføre arbeidet i et nytt prosjekt. Prosjektet vil søke støtte gjennom en ny søknad til Arkivverket, samt søknader til Norges forskningsråd og internasjonale ordninger (bl.a. EØS-midler).

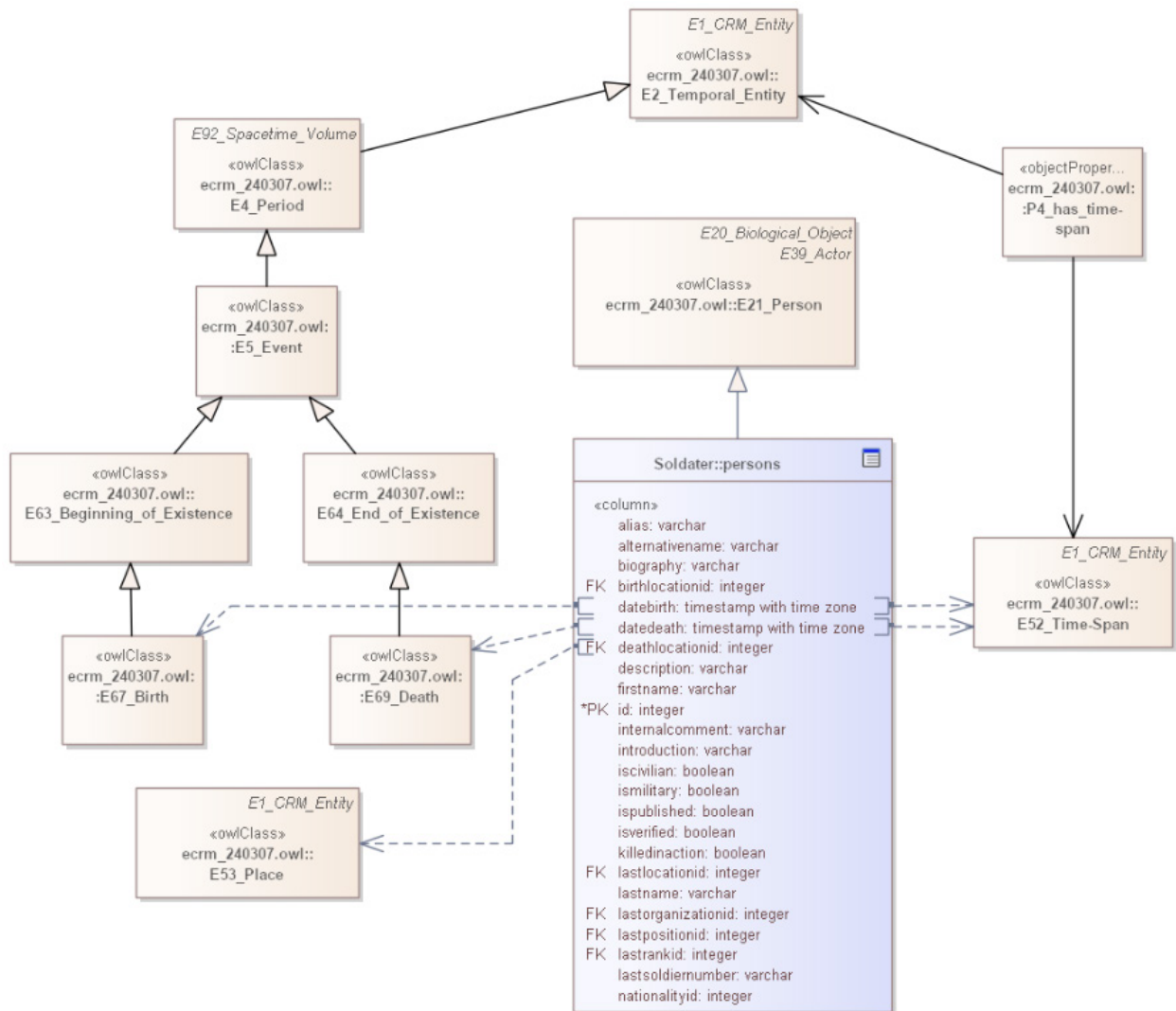
Uavhengig av annen finansiering vil videre arbeid med EDISH fortsette opp mot EHRI-ERIC. Det er også opprettet kontakt med NorwAI, en SFI finansiert av Norges forskningsråd. I videre arbeid ønskes også tettere dialog med Arkivverket og Norsk regnesentral, samt våre assosierte partnere.

Ett av de første stegene i neste prosjekt være å ta tak i de kjente mangler ved konseptmodellen identifisert i EDISH R3 (avsnitt 2.3.4) og i erfaringene fra validering med instanser oppsummert i avsnitt 3. Konseptmodellen vil også utvides til å støtte beskrivelse av mer komplekse hendelsesforløp, inkludert ulike enkelthendelser som henger sammen (og dels avhenger av hverandre).

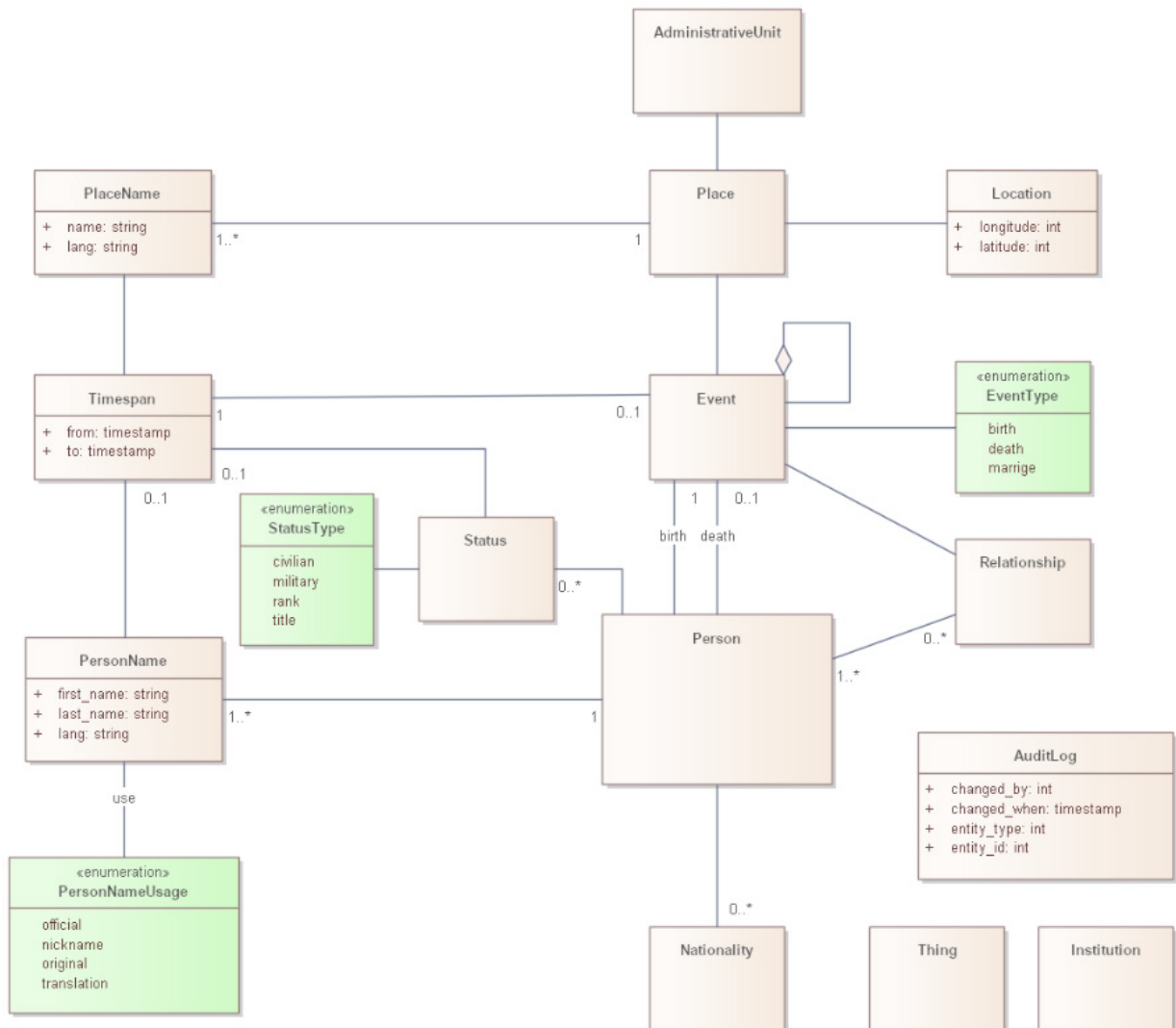
Når den konseptuelle modellen er oppdatert og en mer komplett datamodell er på plass basert på denne, åpner det for andre former for validering av datamodellen. Det kan da for eksempel være aktuelt installere kopier av de opprinnelige databasene samt å opprette en database for den nye datamodellen i et eksperimentelt miljø for validering. I et slikt miljø vil man kunne eksperimentere med integrasjon av informasjon for utvalgte utfordrende eksempler fra de kopierte databasene inn i den nye databasen. Brukerhistoriene fra R2 (avsnitt 2.2) vil være nyttige for å definere hva man validerer og for å velge eksempler i en slik sammenheng.

Ekne 17.11.2025

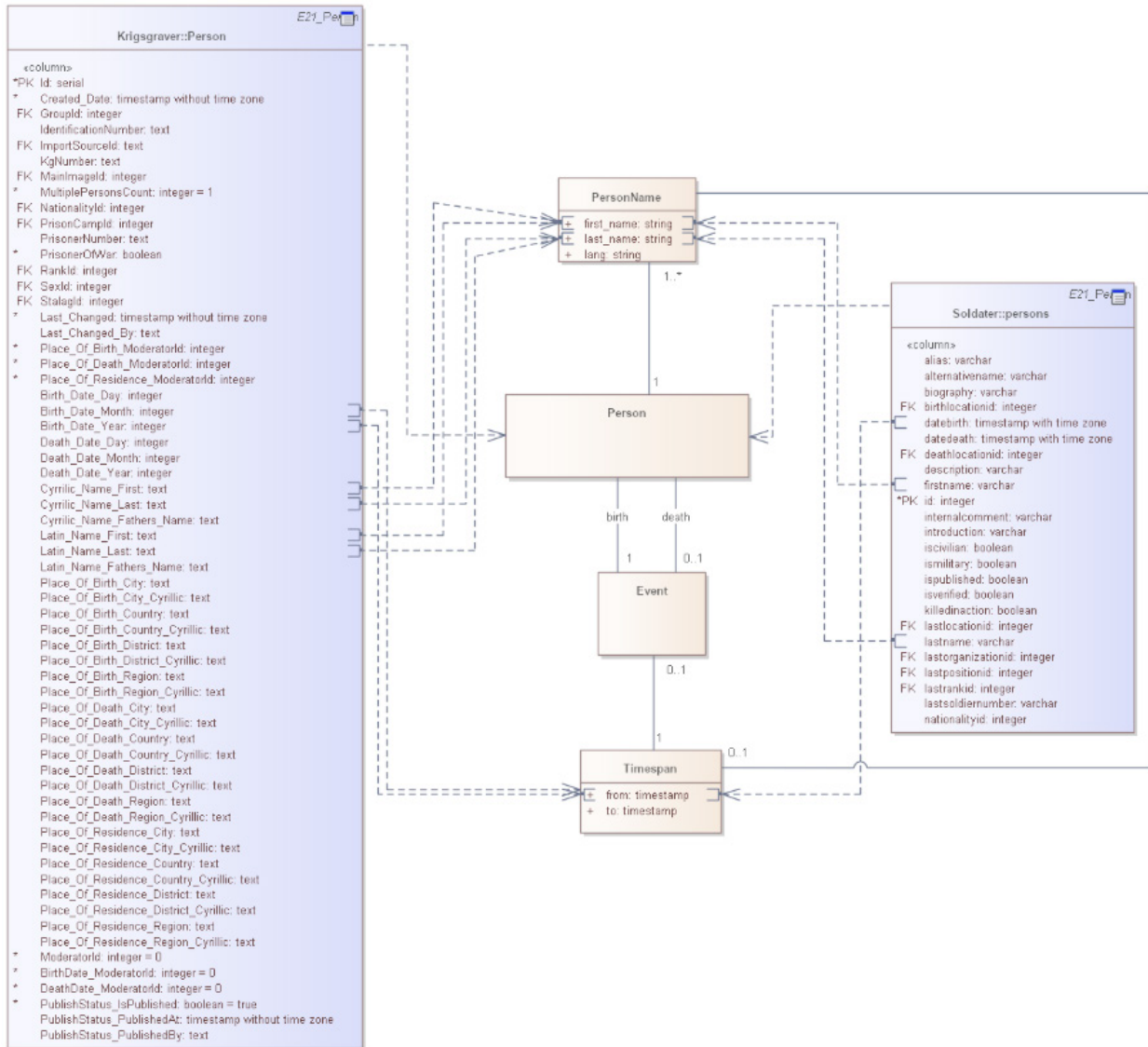
Arne Langås
Prosjektleder, Falstadsenteret



Figur 2. Ontologier. Mapping av fødselsdato, dødsdato og dødssted fra datamodellen i Soldater til konsepter fra CIDOC: hendelsen «Death» og dets «Timespan» som det arver fra «Temporal_Entity». Se avsnitt 2.3.3.



Figur 3: Skisse til konseptuell modell. Skissen håndterer noen av de sentrale problemstillingene i EDISH, inklusive håndtering av person og stedsnavn med historikk over tid. Se avsnitt 2.3.4.



Figur 5: Mapping mellom konseptuell modell og datamodeller fra eksisterende databaser. Se avsnitt 2.4.2.