



LUNDS
UNIVERSITET

LUs flaggskepp: Bevarade och tillgängliga databaser

ALICE OLSSON & SNORRI BRIEM, UNIVERSITETSBIBLIOTEKET VID LU
WORKSHOP SND NÄTVERKSTRÄFF 2024-10-17



- 
- Presentation projekt Bevarade och tillgängliga databaser
 - Fikapaus
 - Hands on-test av verktyget Heurist
 - Uppsamling med tid för frågor

Några begreppsdefinitioner

- (Digital) humaniora och samhällsvetenskap – (D)HS.
- Databas definieras i det här sammanhanget som:
”en webbaserad, interaktiv och relationell IT-lösning för forskningsdata som används och fortsatt uppdateras med data”.¹
- Bevarande – ska här förstås mot bakgrund av Universitetsbibliotekets uppdrag att bevara och tillgängliggöra tryckt material ”[...] för samtid och framtid.”²

¹ Lunds universitet (2022). *Projektdirektiv Bevarade och tillgängliga databaser*. Dnr STYR 2022/309 (2022-04-27).

² Lunds universitet (2019). *Uppdrag till Universitetsbiblioteket*. Dnr STYR 2019/1532 (2019-09-26).

Bakgrund - komplexa forskningsdata inom HS området

Diversifierade praktiker

- Lärosätsgemensamma, ämnesspecifika lösningar saknas trots nationella infrastrukturer, labb/centrumbildningar som erbjuder stöd riktat mot DHS.
- Utveckling av både system och databaser sker på lokal nivå, inte sällan kopplat till lokala projekt.
- Utmaningar att bevara och förvalta databaser långsiktigt.

Kostsamt och resurskrävande

- Tendens att relativt resursintensiva satsningar fått dominera.
- Finansiering genom olika kombinationer av direkta anslag från forskningsfinansiärer, medel från lärosäte/fakultet/institution samt projektmedel.



Projekt BTD – Syfte och mål

Syftet med projektet är att utforma, testa och erbjuda en process för att bevara och tillgängliggöra databaser som nyttjas och uppdateras och som samtidigt bör bevaras långsiktigt. I projektet definieras databas som en webbaserad, interaktiv och relationell IT-lösning för forskningsdata, som används och fortsatt uppdateras med data.

Effektmål

- Erbjuder forskare som producerar databaser att hitta IT-lösningar som stödjer en kombination av tillgängliggörande och bevarande.
- Möjliggör tillgänglighet till levande databaser under en längre period än vad som är möjligt idag.
- Svara mot olika intressenters krav på tillgänglig och FAIR data.



Heurist – plattform för databasutveckling

- Webbaserat verktyg (Open Source) för att utveckla, dela och bevara mindre forskningsdatabaser, skapat med bas i University of Sydney.
- Syftet: möta vad som uppfattas som unika behov av forskningsdatahantering inom humaniora – forskningsdata som är heterogen och med hög grad av komplexitet och relationer.
- Utvecklingsteamet var fram till våren 2024 kopplat till University of Sydneys Faculty of Arts and Social Sciences men har nu ersatts av den ideella föreningen The Heurist Network Association, registrerad i Frankrike.
- Öppna installationer: [Nectar Reserach Cloud/Intersect](#) (Australien), [HumaNum](#) (Frankrike).

<https://heuristnetwork.org/>

Applicerbarhet projekt BTD

- Forskare är främsta målgrupp och stor vikt läggs vid användarvänlighet
- Utvecklingsteam består av en grupp ämnesexperter inom DHS-området.
- Iterativ utveckling möjlig även i senare skede av projekt.
- Stort fokus på export- och importfunktioner och migrering till andra plattformar.
- Möjligheter till samarbete kring utveckling och standarder nationellt och internationellt:
 - Open Source (förutsättning för samarbete kring utveckling och drift)
 - Allt mer decentraliserad utveckling. Centrala utvecklingsteam samordnar flera aktiviteter för ökat samarbete mellan användare.





Resultat från pilot –
Användarbehov

Forskningsprojekt 1

- Nyligen uppstartat forskningsprojekt inom globalhistoria letar efter lämplig IT-lösning för att utveckla en databas.
- Projektgrupp bestående av forskare både från svenska lärosäten och från utlandet. Dessutom har en praktiserande student fått huvudansvar för datainmatning.
- Forskarna söker en plattform där de kan:
 - utveckla och strukturera databasen tillsammans i projektgruppen.
 - sköta all datainmatning via plattformen.
 - tillgängliggöra stora delar av databasen i ett publikt gränssnitt.
- Källmaterial som delvis är komplext och svårtolkat. Forskarna söker sätt kunna markera osäkra eller tvetydiga data och kommentera tolkningar av materialet i samband med datainmatning.
- Projektgruppen har för avsikt att använda plattformen under hela projektiden.



Forskningsprojekt 1

Systemet måste (a) tillåta samarbete i att utveckla databaser och fylla dem med forskningsdata. Samarbete ska kunna ske med forskare från andra lärosäten, studenter eller praktikanter, där alla användare inte nödvändigtvis ska ha samma behörigheter i systemet. Systemet måste (b) tillåta komplexa, osäkra eller tvetydiga forskningsdata, gärna med möjlighet att föra en intern dialog om forskningsdata. Systemet ska även (c) kunna användas under alla steg i utvecklings- och tillgängliggörandeprocessen av forskningsdata.

Samarbete och behörigheter

- Webbaserad plattform som är byggd för samarbete och för att flera användare ska kunna arbeta i databaser samtidigt, både i att utveckla och redigera databasens struktur och inmatning av data.
- Användare kan ha olika behörighetsroller, vilket möjliggör att detaljstyra olika användares behörigheter att se och redigera forskningsdata.
- En databas kan ha flera administratörer, som kan komma åt alla funktioner för att redigera databasens struktur och data. Utöver detta kan arbetsgrupper med begränsad tillgång skapas.



Komplexa och osäkra forskningsdata

- Eftersom systemet är utvecklat med humanioraforskning i åtanke finns det stora möjligheter att arbeta med tvetydiga eller osäkra data.
- Systemet tillåter vissa funktioner för att kommentera datainmatning, där individuella kommentarer kan läggas på varje post.

The screenshot shows a web interface for entering dates, with the 'Fuzzy Range' tab selected. The interface includes a timeline diagram with markers for TPQ, PB, PE, and TAQ, and a 'likely range' shaded area. Below the diagram are input fields for Terminus Post Quem (-3350), Probable begin (-3275), Probable End (-3230), and Terminus Ante Quem (-3105). There are also dropdown menus for Start Profile and End Profile, both set to 'Flat'. The 'Type of Determination' section has radio buttons for Unknown (selected), Attested, Conjecture, and Measurement. At the bottom, there is a 'Use calendar' dropdown set to 'Gregorian' and a 'Comment/original information' text area. A note at the bottom states: 'Temporal objects are converted to a standard format. This field is stored in the temporal to record the original values entered.'

Simple Date | Simple Range | **Fuzzy Range** | Radiometric

Note: Date fields record a single date estimation. They do NOT record a date range. For that you should use Start date and End date fields

TPQ PB PE TAQ

likely range

-3350 Terminus Post Quem set => -3275 Probable begin — -3230 Probable End <= set -3105 Terminus Ante Quem

Start Profile: Flat End Profile: Flat

Entering dates
Year (eg. 1624) or ISO date (yyyy-mm or yyyy-mm-dd)
Use minus (-) for BCE dates (eg. -375 for 375 BCE)

Start & End profile
What is the distribution of the likelihood of the date between the terminus date and the probable date (if in doubt, choose flat, that is any point is equally likely).

Type of Determination
How the dates arrived at?

Unknown
 Attested
 Conjecture
 Measurement

SAVE Cancel

Use calendar: Gregorian

Comment/original information:

Temporal objects are converted to a standard format.
This field is stored in the temporal to record the original values entered.



Verktyg för hela utvecklings- och tillgängliggörandeprocess

- Systemet är byggt för hela utveckling- och tillgängliggörandeprocessen av databaser, dvs. från de första stegen i att strukturera forskningsdata till webbpublicering av hela databaser.
- Att forskningsprojekt använder systemet under hela projekttiden (med möjlighet att bevara databasen efter avslutad projektid) är en av de främsta funktionerna med plattformen.
- Export- och importfunktioner möjliggöra att migrera och byta plattform.



Forskningsprojekt 2

- Databas som skapats inom ramen för ett nyligen avslutat forskningsprojekt i historia.
- Osäkerhet kring hur länge forskarna skulle ha tillgång till det databasverktyget som använts inom projektet, vilket skulle innebära förlorad tillgång till den aktiva databasen.
- Forskarna vill fortsätta uppdatera och utveckla databasen även efter att projektet är avslutat.
- I samband med att forskningsprojektet avslutas vill forskarna även att databasen tillgängliggörs på ett enkelt sätt där de själva har kontroll över de delar av databasen som ska delas.



Forskningsprojekt 2

Systemet ska (d) tillåta forskarna att ha långsiktig tillgång till den aktiva databasen, för att kunna uppdatera databasens forskningsdata och struktur. Systemet måste även erbjuda (e) ett enkelt sätt att dela och tillgängliggöra forskningsdatabaser, där användaren kan detaljstyra vilka forskningsdata som ska tillgängliggöras i ett publikt gränssnitt.



Långsiktig tillgång till aktiv databas

- MySQL-databaser med identisk, väldokumenterad struktur, vilket innebär att databaser kommer att vara läsbara även utan Heurist.
- Arkivpaketet bestående av en textfil som innehåller en MySQL-databasdump, en textfil som innehåller en XML-rendering av databasinnehållet, uppladdade filer (text, bilder, videor, kalkylblad osv), samt en textbeskrivning av Heurist-databasstrukturen.



Tillgängliggöra forskningsdata

- Tillgängliggörande av hela eller valda delar av databaser via ett webbpubliceringssystem är en av grundfunktionerna. Forskningsdata delas via unika webbsidor där användaren kan detaljstyra vilka forskningsdata som ska vara publika och inte.
- Publika söksidor skapas via en grafisk användargränssnittskomponent som tillåter att bädda in sparade filter. Utöver detta finns funktioner för att visualisera forskningsdata på olika sätt – via kartor, tidslinjer, som en nätverkskarta eller i tabellform.
- Grafiska formgivning av är flexibel och det är möjligt att utforma grafiska mallar som kan integreras i olika databaser.



Mall enligt LUs grafiska profil

website editor ase



[Project background](#) [Aim and scope](#) **Search** [Contacts](#) [Credits](#)

Search

Search the whole database with the general search field or use the faceted search below.



Search

Reset all ↶

Namn

Kön

Kvinna

Titel

select...

Civilstånd

Ånka/Ånkling

128

71

Position: 2549 Carin Andersdotter, ID: 495, Titel: Amma, Verksamhet: ej angiv...

Klassificering

Argumentationslinje Egen erfarenhet

LOCALISATION

Verksamhet ej angiven

Land Sverige

DATING

Tidsuppgift typ Tidigare

REFERENCES

Källa RA E1622_2 (from b.100)

Källtyp Supplik.Supplik: yttersta nöd

Innehåller arbetsbeskrivning Nej

Uppgift om plats för utförande Nej

Uppgift om hanterade material Nej

Bild nr 218

Carin Andersdotter, ID: 495, Titel: ej angiven, Verksamhet: ej angiven, Källa: RA E1622_2 (from b.100)



LUNDS
UNIVERSITET



Support, drift och förvaltning

Stöd till anslutning

Inledande supportmöte för att introducera nya användare och skapa förutsättningar för relativt självständigt bruk av systemet.

Ansvarsfördelning, UB i Lund ansvarar för:

- Hårdvara och drift av systemet
- Uppgraderingar av mjukvaran
- System för regelbunden, automatisk backup
- Teknisk support, inklusive komma igång- stöd vid initial konfiguration av databasen

Förutsättningar och behov skiljer sig åt, exempelvis är det mer resursintensivt att flytta existerande databaser (stöd för nödvändiga anpassningar och förberedelser i systemet, förberedelse av forskningsdata och filer osv.).



Användare

- Installationen förvaltas av Universitetsbiblioteket vid LU och är öppen för forskare både på LU och andra lärosäten.
- Trots att projektet fokuserat på behov av hantering och strukturering av forskningsdata inom HS-området bör lösningen kunna användas för:
 - Interaktiva och relationella databaser med mindre datamängd.
 - Forskningsdata utan känsliga data (persondata osv.).
 - Databaser som ska tillgängliggöras, antingen delvis eller i sin helhet.
 - Aktiva eller levande databaser, det vill säga databaser som uppdateras eller används för forskning.

Förvaltningsorganisation

Systemägare
(UB)

IT-systemägare
(UB IT)

Systemförvaltare
(UB Vet.Kom.)

IT-systemförvaltare
(UB IT)



Förvaltningsorganisation - lärdomar

- Djup förståelse för forskningsprocessen och olika behov av forskningsdatahantering inom framför allt HS-området och hur lösningsmodellen bör användas inom verksamheten för att möta behoven (SF).
- Tydlig kommunikation kring Heurist funktion och användningsområde och vilka databaser som ska använda systemet (SF).
- Utveckling på initiativ av användarcommunity med hög frekvens. Stort fokus på utvärdering av uppgraderingar och funktioner utifrån verksamhetens behov och önskemål (SF, IT-SF).
- Vilja och intresse för att delta både i anpassningar lokalt och bidra till utvecklingen centralt (SF, IT-SF).



Hur säkerställa långsiktigt bevarande?

- Bevarandeuppdrag (bevara och tillgängliggöra för samtid och framtid) som även innefattar digitalt material, system vi förvaltar och digitalt födda arkiv. Men vad innebär bevarandeuppdraget rent konkret?
- Metoder för att bevara digitala system ”i all oändlighet” saknas i nuläget, trots att frågan aktualiseras allt mer.
- Metoder för att säkra bevarande, exempelvis LOCKSS.



Vad löser lösningen?

Diversifierade praktiker och kostsamma lösningar:

- Gemensam plattform och process för utveckling och tillgängliggörande med möjlighet att administrera många databaser på samma instans. Tillåter iterativ utveckling av databaser i hela projektets skede.
- Mindre förvaltningsgrupp och gemensam process för att hålla kostnader låga.
- Hur möta behov av specifika anpassningar i olika projekt?

Krav på FAIR:

- Integration av PID-lösning – utveckling sker på LU och Heurist arbetsgrupp.



Kontakt:

Alice Olsson, bibliotekarie, Universitetsbiblioteket
vid Lunds universitet, avdelningen vetenskaplig
kommunikation

alice.olsson@ub.lu.se





LUNDS
UNIVERSITET