

Länkad data på fältet: från fältdokumentation och rapportering till dokumentationskontinuum

Isto Huvila

Institutionen för ABM, Uppsala universitet
förnamn.efternamn@abm.uu.se

1 Introduktion

Arkeologisk dokumentation och rapportering är en process i flera steg. På fältet fångar man data och gör anteckningar som renskrivs och tolkas efteråt, och slutligen sammanfattas i en rapport. Resultaten syntetiseras vidare och publiceras ytterst sällan och det publiceringen kan ske årtal efter den själva utgrävningen. Därtill är det typiskt att olika fyndgrupper och mätningar analyseras och publiceras separat.

Att dokumentationsprocessen är okontinuerlig har vissa nackdelar. Dokumentationsdatan blir tillgänglig långt efter fältarbetet har tagit slut och sen syntetisering av resultaten kan påverka tolkningarnas kvalitet på ett negativt sätt. Dokumentationsprocessen och dess form beror på flera olika faktorer. Tidsbrist på fältet, datamängden, tiden för processering av mätningsdata och analyser, svårigheten att tolka partiella observationer men även traditioner, benägenhet att betrakta primärdata som utgrävningsledarens privata egendom och tendens att skydda egna tolkningar från kritik inverkar på dokumentationsprocessen.

Trots de båda sociala och tekniska svårigheter, det finns en del fördelar med att bearbeta dokumentationsprocessen så att data syntetiseras och tillgängliggörs tidigare och att olika typer av data inte behandlas separat. Denna presentation diskuterar utmaningar och möjligheter vid en integrerad Linked Data (länkad data) baserad approach som gestaltar dokumentation och rapportering som ett kontinuum i stället av en serie (semi)-självständiga faser. Approachen bygger på ett integrerat dokumentationssystem och dataarkiv, och exploatering av semantiska webbt teknologier.

2 Arkeologisk dokumentation på informationskontinuum

Komplexiteten att hantera arkeologisk information har diskuterats sedan länge (Reilly and Rahtz, 1992). Datorisering har underlättat hantering av datamängder och nya mätnings- och analyseringsinstrument har ökat precisionen vid den fångade data. De nya teknologierna har dock inte gett arkeologerna nya revolutionerande verktyg eller lösningar för att hantera den fundamentala komplexiteten av arkeologisk information. Tvärtom de nya teknologierna har ökad mängden data och understrukt behovet av effektivt hantera data (Lock, 2003). Samtidigt, samhällseliga förväntningar av ett ökat samarbete och påtagliga möjligheter och fördelar vid större

integrerade databaser och e-vetenskapsinfrastrukturer har uppmuntrat arkeologerna att utveckla metoder för en effektivare och mera ändamålsenlig hantering av primärdata.

Den första uppenbara utmaningen handlar om komplexiteten vid olika typer av data och information. Ett relaterat problem är att olika typer av data och datakällor baserar sig på olika nivåer av precision och kvalitet. Den slutliga tolkningen av undersökningsdata, den kvalitativa förståelsen av det förflutna, är ett resultat av en komplex tolkning av en blandning kvalitativ och kvantitativ data från journalanteckningar, mätningar, naturvetenskapliga analyser, tolkning av föremål, sampel, ritningar och, till exempel, foton (Greene, 1998). De olika typerna av material är sällan jämförbara i sig och hanterbara inom ett enhetligt system. Den andra utmaningen berör ofullständighet av de befintliga data. Arkeologer har tillgång endast till ett litet slumpmässigt urval av de fullständiga data som funnits. Konsekvensen av de "bristfälliga" data är att relationen mellan data och slutliga tolkningar är sällan entydig (Thomas, 2006).

Den tredje, informationshanteringsrelaterade utmaningen är att en typisk arkeologisk informationsprocess har betydliga diskontinuiteter (Greene, 1998). På en typisk arkeologisk undersökning en individ eller grupp har ansvar för mätningar. En skild grupp gräver fyndplatsen, samlar in fynd och dokumenterar sina observationer på standardiserade formulär. Det är möjligt att en tredje grupp hanterar fynd och vissa andra personer ta alla foton på fyndplatsen. Även om en person skulle arbeta med olika typer av data och datainsamling, i slutskede de olika typerna av data hamnar ofta i olika system. Ofta endast undersökningsledaren och assistenten har ansvar och möjligheter för att skapa en allmän uppfattning om helhet på fyndplatsen. I praktiken finns dock endast begränsade möjligheter att syntetisera data redan på fältet. I vanliga fall den rapporterade helhetstolkningen börjar forma sig först långt efter fältundersökningen utifrån anteckningar, fältdata och minnesbilder som kopplar ihop olika enskilda typer av data. Hantering och analys av individuella typer av data utgörs oftast av enskilda experter eller grupper. En del av arbetet kan anlitas på individer som inte har tagit del i utgrävningen eller har varit närvarande på fältet bara en del av tiden. Separat behandling av data fortsätter i rapporter och till och med publikationer (om resultat publiceras överhuvudtaget) som ofta presenterar olika typer av fynd och data i enskilda kapitel.

Thomas (2006, 30) understryker att diskontinuiteter i dokumentationsprocessen är ett betydande problem. Det typiska stegvisa arbetssättet är att det kan leda till betydliga klyftor mellan datafångst, analys och rapportering. Dessa klyftor leder till luckor i informationskontinuum (Oliver, 2010), i den processen hur olika typer av forntida och historiska materiella föremål (information) förvandlas till kvarlevor som grävs upp, analyseras och används (information) vid vetenskapliga tolkningar (information) som omformas till populärvetenskapliga bilder (information) av det förflutna. Ett brutet informationskontinuum ökar sannolikheten av tolkningar av onödigt dålig kvalitet. Ett vidare problem med diskontinuiteten är att klyftor är svåra att överbygga efteråt. Den arkeologiska informationsprocessen baserar sig på kumulativa tolkningar (Thomas, 2006, 30) vilket försvårar möjligheter att efteråt åtgärda diskontinuiteter.

3 Länkad data på fältet

Ett möjligt sätt att undvika klyftor i informationskontinuum är att försöka hantera allt information på ett enhetligt sätt på ett ställe under hela dokumentationsprocessen. Detta testades på två olika undersökningar i Egentliga Finland tillsammans med arkeolog, docent Kari Uotila. Forskningsmetoden baserade sig på aktionsforskning (Greenwood and Levin, 2000; Kemmis and McTaggart, 2000). Den tekniska lösningen baserade sig på en semantisk wiki, ett wikisystem med tilläggsfunktioner för semantisk markering av data (?). Systemet hade testats förut i samband med hantering av information om ett gårdsarkiv och två medeltida slott (Huvila, 2008).

Grundtanken bakom systemet var att alla stratigrafiska enheter, fynd och observationer ska dokumenteras i informationshanteringssystemet direkt på fältet och alla 'saker' ska ges omedelbart ett permanent identifikation som samtidigt är en URL-adress. I praktiken alla dokumenterade entiteter på undersökningen blir *länkad data* [Needleman (2011)]. Syftet med approachen är att allt data hanteras på ett centralt ställe, det är lätt att hänvisa till andra 'saker' inom systemet och till och med utanför systemet på den öppna webben. Därtill är det lätt att utforska vad andra medlemmar i forskningsgruppen gör och dokumenterar och utforma tolkningar om materialet redan i ett tidigt skede. Den tekniska lösningen, ett wikisystem, tar hand om versionshantering (alla ändringar dokumenteras), den fungerar väl i kollaborativt arbete och kan hantera olika typer av data.

De främsta fördelar med approachen är att det är lätt att kombinera formell och fritt formulerad dokumentation. Det är möjligt att beskriva fynd genom att använda formella kategorier och samtidigt lägga till preliminära observationer och fria beskrivningar i text och bild. Fördelen med ett enhetligt system är att det är befintlig dokumentation är lätt att komplettera och utvidga med nya beskrivningar, kategoriseringar, analysresultat och referenser. De största praktiska utmaningarna handlar om konceptualisering av betydelsen av länkar mellan olika entiter, hantering av parallella och motstridiga tolkningar, hanterbarheten av ett relativt öppet system och samordning av det arkeologiska och wikibaserade arbetssättet.

Utöver närmast tekniska utmaningar, en länkad data approach has vissa implikationer i den arkeologiska dokumentationsprocessen på en mera teoretisk nivå. Wikisystem använder typisk läsbara URL-adresser i stället av ett ID-nummer. URL-adressen innehåller klassifikatorisk information samt ett namn på entiteten. Systemet har uppenbara fördelar. Det är lättare att skapa länkar mellan olika entiter genom att använda tydliga namn och det är även lättare att få en överblick på entiteten genom att läsa URL-adressen. Ett liknande systematik är typisk även inom befintliga korpusar av länkad arkeologisk data. Hittills den största delen av projekt med syfte på att skapa länkad data inom arkeologi har dock fokuserat på befintliga databaser med 'färdiga' strukturer och tolkningar (e.g. Tudhope et al., 2011). Skapandet av meningsfulla adresser på fältet medan tolkningar och relationer mellan olika entiteter fortlöpande förändras, entiteter skapas, slås ihop och tas bort, är svårt. En möjlighet är att använda id-numrering med läsbara alias och ett effektivt system för hantering av adresser, alias och samband mellan entiteter.

En annan utmaning handlar om tolkningsrelaterade konsekvenser av den valda approachen. Medan den praktiska utmaningen är att lära arkeologer använda ett system som baserar sig på

länkad data, den teoretiska utmaningen är att förstå hur systemet påverkar den arkeologiska intellektuella processen. Hur länkad data kommer att förändra arkeologi? En annan lika fundamental fråga handlar om semantiken på den semantiska webben och länkad data och den semantiken i arkeologi. Almeida, Souza och Fonseca (2011) har skrivit en kritisk granskning på den typen av semantiken som fungerar som premisser för den semantiska webben. Författarna konkluderar att det finns en betydlig diskrepans mellan den formella semantiken på den semantiska webben och hur semantiken förstås, till exempel, inom humanistiska vetenskaper.

Sammanfattningsvis kan konstateras att det finns flera fördelar att försöka konsolidera arkeologiska dokumentationsprocesser så att klyftor i informationskontinuum kan undvikas. Samtidigt är det viktigt att stå kritisk till följd effekter av semantificering och överväga noggrant hur olika tekniska lösningar påverkar inte bara det praktiska arbetet utan även den arkeologiska tolkningsprocessen.

References

- Almeida, Mauricio, Renato Souza, and Fred Fonseca. 2011. Semantics in the semantic web: A critical evaluation. *Knowledge Organization* 38(3):187–203.
- Greene, Kevin. 1998. *Archaeology: An introduction*. London: Routledge.
- Greenwood, Davydd J., and Morten Levin. 2000. Reconstructing the relationships between universities and society through action research. In *Handbook of qualitative research*, ed. Norman K. Denzin and Yvonna S. Lincoln, 85–106. Thousand Oaks, CA: Sage.
- Huvila, Isto. 2008. Participatory archive: towards decentralised curation, radical user orientation and broader contextualisation of records management. *Archival Science* 8(1):15–36.
- Kemmis, Stephen, and Robert McTaggart. 2000. Participatory action research. In *Handbook of qualitative research*, ed. Norman K. Denzin and Yvonna S. Lincoln, 567–605. Thousand Oaks, CA: Sage.
- Lock, Gary. 2003. *Using computers in archaeology: Towards virtual pasts*. London: Routledge.
- Needleman, Mark H. 2011. Linked data: What is it and what can it do? *Serials Review* 37(3):234.
- Oliver, Gillian. 2010. Transcending silos, developing synergies: libraries and archives. In *Information research 15 (4). special supplement: Proceedings of the Seventh International Conference on Conceptions of Library and Information Science – Unity in diversity – Part 2*.
- Reilly, Paul, and Sebastian Rahtz. 1992. *Archaeology and the information age*. London and New York: Routledge.
- Thomas, Julian. 2006. The great dark book: Archaeology, experience, and interpretation. In *A companion to archaeology*, ed. John Bintliff, 21–36. Malden and Oxford: Blackwell Publishing.
- Tudhope, Douglas, Ceri Binding, Stuart Jeffrey, Keith May, and Andreas Vlachidis. 2011. A STELLAR role for knowledge organization systems in digital archaeology. *Bulletin of ASIS&T* 37(4):15–18.