

VAATENURK

KUST PÄRINEVAD METAANDMED: INFOTEOREETILINE VAADE

Kurmo Konsa

Metaandmetega puutub tänapäeva inimene kokku pea iga päev. Otsides midagi veebist, kasutame märksõnu ja veebilehekülgede aadresse; vaadates raamatupoes uusi raamatuid, köidab meie tähelepanu teose pealkiri ja autor; selleks et saata sõbrale e-kiri, peame teadma tema e-posti aadressi. Kõikide nende olukordade puhul on meil tegemist metaandmetega. Kuigi metaandmete olulisus on pärast digiinfo kasutuselevõttu märgatavalt kasvanud, on neid kasutatud juba alates esimeste mäluasutuste loomisest.

Metaandmetega seotud uuringud keskenduvad peaaegu ainult praktiliste küsimuste lahendamisele.¹ Vähesed metaandmete teooriat käsitlevad uurimused rõhutavad, et puudub ühtne teoreetiline alus ning et vastavaid töid napib.² Metaandmete loomisel ja haldamisel kasutatavad lahendused põhinevad ainult praktilistel kogemustel, mitte teoreetilistel lähenemistel.³ Jane Greenberg toob esile ka kaks asjaolu, mis on sellise olukorra põhjustanud. Esiteks on metaandmete mõiste ise liiga uus ning teiseks pole lihtsalt

¹ Ülevaadet vt Miguel-Angel Sicilia, "Metadata research: making digital resources useful again", *Handbook of metadata, semantics and ontologies*, ed. by Miguel-Angel Sicilia (Singapore: World Scientific Publishing Co. Pte. Ltd, 2014), 1–8.

² Vt nt Jane Greenberg, "Theoretical Considerations of Lifecycle Modeling: An Analysis of the Dryad Repository Demonstrating Automatic Metadata Propagation, Inheritance, and Value System Adoption", *Cataloging & Classification Quarterly*, 47 (2009), 380–402.

³ Jane Greenberg, "Big Metadata, Smart Metadata, and Metadata Capital: Toward Greater Synergy Between Data Science and Metadata", *Journal of Data and Information Science*, 2:3 (2017), <https://doi.org/10.1515/jdis-2017-0012>.

spetsialiste, kellel oleks aega tegeleda teooriaga.⁴ Veelgi kaalukam on aga asjaolu, et metaandmeid peetaksegi pelgalt tehnilisse valdkonda kuuluvaks. Sellist lähenemist rõhutab näiteks USA Riiklik Infostandardite Organisatsioon oma metaandmeid käsitlevas juhtdokumendis.⁵ Teoreetilisemat laadi on metaandmete ajaloo käsitlused, mis keskenduvad interneti metaandmete seostamisele varasemate teavikute kirjeldussüsteemidega.⁶ Eelkõige tegeletakse metaandmete klassifitseerimise ja nende funktsioonide määratlemisega, kuid need tööd on pigem praktilise suunitlusega.⁷ Infoteoreetilisemat laadi käsitlusi on ilmunud metaandmete väärtustest.⁸ Süsteemiteoreetilist lähenemist on kasutatud üht tüüpi metaandmete, nimelt ontoloogiate analüüsimiseks.⁹ Metaandmete loomist on uuritud sotsiaalkonstruktivistliku filosoofia vaatenurgast.¹⁰ Laiemasse teoreetilisse konteksti püüab metaandmeid asetada uurimus, mille autorid seostavad metaandmete kasutamise arengu Darwini evolutsiooniteooria põhipostulaatidega.¹¹

Siinne artikkel vaatlleb metaandmeid laias infoteoreetilises võtmes ning otsib vastust esmapilgul lihtsale küsimusele – kust tulevad metaandmed? Sellele küsimusele vastuse leidmiseks annan kõigepealt lühikese ülevaate

⁴ Greenberg, “Theoretical Considerations of Lifecycle Modeling”, 380–402, 383–384.

⁵ Jenn Riley, *Understanding Metadata: What is metadata, and what is it for* (Baltimore: National Information Standards Organization, 2017).

⁶ Vt nt Lynne C. Howarth, “Metadata and Bibliographic Control: Soul-Mates or Two Solitudes?”, *Cataloging & Classification Quarterly*, 40:3 (2005), 37–56; Jane Greenberg, “Understanding Metadata and Metadata Schemes”, *Cataloging & Classification Quarterly*, 40:3/4 (2005), 17–36.

⁷ Vt nt Marieke Guy, Andy Powell, and Michael Day, “Improving the Quality of Metadata in Eprint Archives”, *Ariadne* 28 (2004), <http://www.ariadne.ac.uk/issue28/guy>; Jane Greenberg, “Metadata and the World Wide Web”, *Encyclopedia of Library and Information Science*, ed. by Marcia J. Bates, Mary Niles Maack, Miriam Drake (New York: Marcel Dekker, Inc., 2003), 1876–1888; Anne J. Gilliland, “Setting the Stage”, *Introduction to Metadata*, Version 3.0, ed. by Murtha Baca (Los Angeles, CA: Getty Information Institute, 2008), http://www.getty.edu/research/conducting_research/standards/intrometadata/setting.html; Greenberg, “Understanding Metadata and Metadata Schemes”, 17–36.

⁸ Miltiadis D. Lytras, Miguel-Ángel Sicilia, Cristian Cechinel, “The value and cost of metadata”, *Handbook of metadata, semantics and ontologies*, ed. by Miguel-Ángel Sicilia (Singapore: World Scientific Publishing Co. Pte. Ltd, 2014), 41–62.

⁹ Emilia Currás, “Ontologies in systems theory”, *Handbook of metadata, semantics and ontologies*, ed. by Miguel-Ángel Sicilia (Singapore: World Scientific Publishing Co. Pte. Ltd, 2014), 89–107.

¹⁰ Getaneh Alemu, Brett Stevens, Penny Ross, “Towards a conceptual framework for user-driven semantic metadata interoperability in digital libraries: A social constructivist approach”, *New Library World*, 113:1/2 (2012), 38–54.

¹¹ Erik Mannens, Ruben Verborgh, Seth van Hooland, Laurence Hauttekeete, Tom Evens, Sam Coppens, Rik van de Walle, “On the origin of metadata”, *Information*, 3 (2012), 790–808.

sellest, mis on metaandmed ning seejärel seostan need kommunikatsiooni-protsessidega. Artikli eesmärk on esitada kontseptuaalne mudel, mis seostab metaandmed kommunikatsiooniprotsessidega ning loob seeläbi võimaluse metaandmete süstematilisemaks käsitlemiseks.

Mis on metaandmed?

Kõige lihtsama ja levinuma määratluse kohaselt on metaandmed andmed andmete kohta ehk andmed, mis kirjeldavad mingeid teisi andmeid. Samas ei ütle see määratlus midagi selle kohta, milleks on üldse vaja andmeid andmete kohta ning mida metaandmed peaksid kajastama.

Kirjandusest võib leida terve rea metaandmete määratlusi. Toon siinkohal mõned neist:

- metaandmed on informatsioon, mis muudab andmed kasutatavaks;¹²
- metaandmed kirjeldavad ressursside erinevaid tunnuseid;¹³
- metaandmed kirjeldavad piiritletud andmeobjekte;¹⁴
- metaandmed pakuvad dokumentide kasutajale otstarbekat teavet dokumentide olemasolu ja tunnuste kohta;¹⁵
- metaandmed kirjeldavad mingi informatsiooni sisu, vormingut ja/või tunnuseid;¹⁶
- metaandmed on struktureeritud informatsioon, mis kirjeldab, selgitab ja/või näitab infoobjekti asukohta või aitab mõnel muul viisil teavet leida, kasutada ja hallata;¹⁷

¹² Martin Grottschel, Joachim Lugger, "Scientific Information System and Metadata", *Classification in the Information Age. Studies in Classification, Data Analysis, and Knowledge Organization*, ed. by Wolfgang Gaul, Hermann Locarek-Junge (Berlin, Heidelberg: Springer Verlag, 1999), 3–20.

¹³ Lorcan Dempsey, Rachel Heery, Martin Hamilton, Debra Hiom, Jon Knight, Traugott Koch, Marianne Peereboom, Andy Powell, "A review of metadata: a survey of current resource description Formats", *Specification for resource description methods* (Vol. 1), ed. by Lorcan Dempsey, Rachel Heery (University of Bath, UK Office for Library and Information Networking, 1997).

¹⁴ Tony Gill, "Metadata and the World Wide Web", *Introduction to metadata: pathways to digital information*, ed. by Murtha Baca (Los Angeles: Getty Research Institute, 2008), 20–37.

¹⁵ Lorcan Dempsey, Rachel Heery, "Metadata: a current view of practice and issues", *Journal of Documentation*, 54:2 (1998), 145–172.

¹⁶ David Haynes, *Metadata for information management and retrieval* (London: Facet Publishing, 2004).

¹⁷ Riley, *Understanding Metadata*.

- metaandmed on andmed, mis kirjeldavad teavikute konteksti, sisu ja struktuuri ning nende haldusprotsessi.¹⁸

Kokkuvõtlikult saab öelda, et metaandmed kirjeldavad mingisuguse infoobjekti tunnuseid, mis on vajalikud selle infoobjekti leidmiseks, kasutamiseks, mõistmiseks, haldamiseks ja säilitamiseks. Infoobjektiks on siinkohal piiritletud üksusena käsitletav informatsioon, mida suudab kasutada inimene või infosüsteem. Tegemist võib olla ühe objektiga, näiteks raamatu või failiga, aga ka objektide kogumi või terve arhiivi, raamatukogu või andmebaasiga. Metaandmeid kasutavad seega nii inimesed kui ka arvutiprogrammid eri ülesannete täitmisel. Infoobjekti ja metaandmete vaheline piir ei ole jäik. Ühe infoobjekti metaandmed võivad samal ajal olla teise infoobjekti andmeteks. Suurel määral sõltub see infoobjektide ja -süsteemide vahelistest seostest ning sellest, millisel eesmärgil kasutaja seda infot käsitleb.¹⁹

Sellises laias tähenduses ulatub metaandmete kasutamine tagasi juba esimeste arhiivide juurde. Teadaolevad esimesed arhiivid on leitud Lähis-Idas ja Mesopotaamia aladel asuvatest linnadest ning need pärinevad ajajärgust 3000–2000 aastat eKr. Varastest arhiividest ei ole leitud arhiivimistuid, -katalooge ega -kirjeldusi. Tõenäoliselt ei koostanud arhiivide eest vastutavad ametnikud kirjeldavaid metaandmeid. See ei tähenda aga seda, et metaandmed üldse puudusid. Dokumentide haldamiseks kasutati nende kindlat asetust hoidlas, eri dokumendivorme ja kronoloogilist järjestust. Nii olid näiteks Ebla linnas asuvas administratiivdokumente sisaldavas arhiivis säilikud organiseeritud nende funktsiooni või teema järgi ning seejärel kronoloogiliselt kuude kaupa.²⁰ Arhiivaalide nimekirjad võeti kasutusele alles 12.–15. sajandil Itaalia linnades.²¹

Vana-Egiptuse tuntuimas antiikaja raamatukogus, 3. saj eKr rajatud Aleksandria raamatukogus, olid teosed paigutatud ja kataloogitud Küreene Kallimachose ja Zenodotose loodud süsteemi kohaselt. Zenodotos, kes oli Aleksandria raamatukogu esimene raamatukoguhoidja, jaotas kõik teosed

¹⁸ ISO 15489-1:2016. Information and documentation. – Records management – Part 1: Concepts and principles.

¹⁹ Anne Gilliland-Swetland, "Setting the stage: defining metadata", *Introduction to metadata: pathways to digital information*, ed. by Murtha Baca (Los Angeles: Getty Research Institute, 2008), 1–19.

²⁰ Alfonso Archi, "Archival record-keeping at Ebla, 2400–2350 B.C.", *Ancient archives and archival traditions: concepts of record-keeping in the ancient world*, ed. by M. Brosius (Oxford: Oxford University Press, 2003), 17–36.

²¹ Luciana Durant, "Origin and Development of the Concept of Archival Description", *Archivaria*, 35 (1993), 48–49.

kolme rühma, mis asetsesid eraldi ruumides. Igas rühmas järjestati raamatud alfaabetsiliselt autorite järgi. Et teoseid oleks lihtsam leida, kinnitati iga papüüruserulli külge väljaulatuv silt, millele olid kirjutatud autor, pealkiri ja teema. Tegemist oli esimese tänapäevases mõistes metaandmete kasutamisega inforessursside korrastamiseks. Kallimachos täiendas teoste jagunemissüsteemi ja koostas esimese raamatukogukataloogi (*pinakes*), mis kandis pealkirja "Igas teadusharus väljapaistvate õpetlaste nimekiri koos nende teostega" ning koosnes 120 papüüruserullist. Autorid olid järjestatud alfaabetsiliselt koos lühikese elulookirjeldusega. Kui autorit ja pealkirja polnud võimalik tuvastada, kataloogiti raamat esimeste sõnade järgi. Raamatute arvelevõtmisel pandi kirja ostukoht, eelmise omaniku nimi ning kommentaatorid, toimetajad ja korrektorid. Tegemist oli esimese tõelise mäluasutuses kasutatud kataloogiga.²² Ka tänapäevaste muuseumite eelastutuste, kuriositeetide kabinetide (15. sajandi lõpp kuni 18. sajandi esimene pool) juures koostati kogude nimekirju ja katalooge.²³ Võrreldes aga teiste mäluasutustega (arhiivid ja raamatukogud) on muuseumide dokumenteerimissüsteemid tänase päevani märksa vähem standardiseeritud ja erinevad muuseumides vägagi olulisel määral.²⁴ Museaalideks on harilikult utilitaarsed ja loodusobjektid, mille polüseemiline olemus muudab nende kirjeldamise olulisemalt keerukamaks võrreldes informatsiooniliste artefaktidega (raamatud, dokumendid, filmid jms).²⁵

Metaandmete termin ja selle kasutus

Teadaolevalt kasutas terminit "metaandmed" esimest korda USA arvuti-teadlane Philip Bagley 1968. aastal.²⁶ Sisuliselt, terminit küll otseselt kasutamata, käsitles andmete kirjeldamiseks kasutatavaid andmeid ka arvutikonsultant George H. Mealy aasta varem ilmunud artiklis²⁷ ning samal

²² Heather Phillips, "The Great Library of Alexandria?", *Library Philosophy and Practice* (e-journal, 2010), <https://digitalcommons.unl.edu/libphilprac/417>.

²³ Arthur MacGregor, *Curiosity and Enlightenment: Collectors and Collections from the Sixteenth to the Nineteenth Century* (New Haven and London: Yale University Press, 2007), 60–64.

²⁴ Ross Parry, *Recoding the Museum* (London, New York: Routledge, 2007).

²⁵ Trilce Navarrete, John Mackenzie Owen, "The Museum as Information Space: Metadata and Documentation", *Cultural Heritage in a Changing World*, ed. by Karol Jan Borowiecki, Neil Forbes, Antonella Fresca (Springer, 2016), 111–123.

²⁶ Philip R. Bagley, *Extension of programming language concepts* (Philadelphia: University City Science Center, 1968), 26, <https://apps.dtic.mil/dtic/tr/fulltext/u2/680815.pdf>.

²⁷ George H. Mealy, "Another look at data", *Managing Requirements Knowledge*, International Workshop on Nov. 14 1967 to Nov. 16 1967 Anaheim, Vol 1 (Los Alamitos: IEEE Computer Society, 1967), 525–534.

aastal tegid seda ka David Griffel ja Stuart McIntosh oma projektiaruandes.²⁸ Viimased kasutasid seejuures terminit “metakeel” (*meta language*). Esialgu piirduski selle mõiste kasutusvaldkond andmebaaside halduse ja programmeerimiskeelte käsitlustega. 1990. aastate keskpaigast võeti termin laiemalt kasutusele info- ja raamatukoguvaldkonnas.²⁹ Tänapäeval kasutatakse terminit “metaandmed” peamiselt elektroonilisel kujul oleva info kohta.

Termini käsitlusviisid pärinevadki ühelt poolt arvutiteadusest ja teiselt poolt raamatukogundusest. Seda peetakse ka üheks põhjuseks, miks on metaandmete määratlused üsnagi laialivalguvad ja käsitlused küllaltki erinevad.³⁰

Metaandmed täidavad peamiselt viit olulist funktsiooni.³¹

- Ressursside kirjeldus – metaandmed võimaldavad objekte identifitseerida, kataloogida ning nende konteksti säilitada. Konteksti säilitamine on eriti oluline selliste objektide puhul, mis moodustavad ühe osa mingist kogust või kollektsioonist (arhivaalid arhiivis, raamatud raamatukogus, museaalid muuseumikogus jne).
- Infootsing – otsisüsteemid kasutavad metaandmeid inforessursside otsimisel. Objekti kasutamisega seotud administratiivne info (võimalikud juurdepääsupiirangud, kasutamise ajalugu jms) ja muud metaandmed aitavad kasutajal infot üles leida, kätte saada ja kasutada.
- Inforessursside haldamine – metaandmete kasutamine on üks parimaid viise, kuidas tagada infoobjektide haldamine, sealhulgas ka säilitamine. Infoobjektide säilitamise käigus toimub nende muutmine, metaandmed on vahend, mis tagab inforessursside sisu püsivuse (*persistence*) ning objektide sõltumatuse nende säilitamiseks ja kasutamiseks mõeldud tehnilisest süsteemist.
- Inforessursside omandiõiguse ja autentsuse tagamine – metaandmed sisaldavad infot objekti omandi- ja autoriõiguse, päritolu, struktuuri

²⁸ David Griffel, Stuart McIntosh, *ADMINS – A Progress Report*. Center for International Studies, Massachusetts Institute of Technology, 1967, <https://dspace.mit.edu/bitstream/handle/1721.1/82974/09487802.pdf?sequence=1>.

²⁹ Nick Roussopoulos, Leo Mark, “Metadata management”, *Computer*, 19:12 (1986) 26–36.

³⁰ Kathleen Burnett, Kwong Bor Ng, Soyeon Park, “A Comparison of the Two Traditions of Metadata Development”, *Journal of the American Society for Information Science*, 50:13 (1999), 1209–1217.

³¹ Wendy Duff, Sue McKemmish, “Metadata and ISO 9000 compliance”, *Information Management Journal*, 34:1 (2000); Haynes, *Metadata for Information Management and Retrieval*.

- ja konteksti kohta ning on seega otsustava tähtsusega objekti autentsuse ning struktuurse ja funktsionaalse terviklikkuse hindamisel.
- Koostalitlusvõime ja infovahetuse võimaldamine – metaandmete vormingud võimaldavad andmeid süsteemide vahel vahetada neid seejuures õigesti interpreteerides ja informatsiooni tähendust säilitades. Inforessursi kirjeldamine metaandmetega muudab selle arvutisüsteemidele mõistetavaks ja aitab sellega kaasa koostalitlusvõimele.

Seega on metaandmed vajalikud nii info haldajale kui ka kasutajale. Info lõppkasutajal ei ole tavaliselt tarvis inforessursi metaandmeid näha, need on tähtsad teabe organiseerimise ja haldamise seisukohast. Kuna metaandmed täidavad nii mitmekesiseid funktsioone, siis on selge, et metaandmete süsteeme on väga palju. Samuti saab metaandmete süsteeme kategoriseerida mitmeti. Üks võimalus on jagada metaandmete standardid ja skeemid funktsioonide järgi nelja gruppi.³²

KIRJELDAVAD METAANDMED määravad infoobjekti bibliograafilised tunnused. Tegemist on infoga, mis identifitseerib inforessursse, aitab neid üles leida ja dokumenteerib suhteid eri inforessursside vahel. Kirjeldavad metaandmed võivad lähtuda mõnest kindlast teavikutüübist, näiteks FGDC (Federal Geographic Data Committee) standardist ruumiliste digiandmete kohta või ISAD(G) (*General International Standard Archival Description*) standardist arhivaalide kohta. Teised metaandmete vormingud, näiteks MARC 21 või Dublin Core, on üldisemad ja sobivad eri teavikutüüpide kirjeldamiseks. Kirjeldavad metaandmed võivad sisaldada ka teavet ajaloolise konteksti kohta: näiteks raamatu kohta, mille digimisel digiobjekt on saadud (allika päritolu).

STRUKTURAALSED METAANDMED määravad loogilised või füüsilised suhted infoobjekti eri osade vahel, mis kirjeldavad näiteks peatükkide asetust raamatus, keeruka digiobjekti failidevahelisi seoseid vms. Levinuim standard on METS.

TEHNILISED METAANDMED määravad info, mis kajastab digiobjekti töötlemiseks vajalikku riist- ja tarkvara, ning info kontrollsummade ja digiallkirjade kohta, mis tagavad infoobjekti autentsuse. Selliste standardite näiteks

³² Angela Dappert, Markus Enders, "Digital preservation metadata standards", *ISO Information Standards Quarterly*, 22:2 (2010), 4–13, 6.

võib tuua fotokujutiste standardi MIX (*Metadata for Images in XML*), tekstistandardi TEI (*Text Encoding Initiative*), audiostandardi audioMD (*Audio Technical Metadata Extension Schema*) ja videostandardi videoMD (*Video Technical Metadata Extension Schema*). Tehnilised metaandmed võidakse paigutada ka administratiivsete metaandmete hulka.³³

ADMINISTRATIIVSED METAANDMED määravad info, mis on seotud info-objektiga tehtud toimingute, st loomise, haldamise, säilitamise ja kasutamisega, sealhulgas kõikvõimalikud muudatused ja otsustused, samuti info õiguste ja objektile juurdepääsu kohta. Näiteks intellektuaalomandi õigusi kajastavate metaandmete standard on ODRL (*Open Digital Rights Language*).

Anne Gilliland on pakkunud välja teistsuguse võimaluse metaandmete kategoriseerimiseks, mis põhineb infoobjektide omadustel. Nimelt on igal infoobjektil, sõltumata selle konkreetsest füüsilisest kujust, kolm iseloomulikku tunnust: sisu, kontekst ja struktuur. Metaandmed peavad kajastama kõiki tunnuseid.³⁴ Viimasel ajal on hakatud kõnelema ka paraandmetest (*paradata*), mis kujutavad endast andmeid kindla teaberessursi täpsemast uurimisest saadud teabe kohta. Kui näiteks mingi museaali uurimisest ilmub artikkel, siis paraandmete kaudu seostub info artikli kohta museaali metaandmetega.³⁵ Metaandmete erinevad jaotused on vajalikud nende lihtsamaks käsitlemiseks, tegelikkuses metaandmete eri grupid suuresti kattuvad ja selged piirid nende vahel puuduvad.

Mäluasutustes kasutatavad kirjeldavad metaandmed

Metaandmed luuakse kindla eesmärgi põhjal. Ei ole olemas ühtset metaandmete skeemi või standardit, mis sobiks igasuguste eesmärkide jaoks. Metaandmete standardid on kindlate organisatsioonide poolt koostatud eeskirjad või juhendid, mis on eeskätt mõeldud koostalitlusvõime suurendamiseks. Standardid võivad kasutada või soovitada kasutada kindlaid

³³ Bernard J. Hurley, John Price-Wilkin, Merrilee Proffitt, Howard Besser, *The making of America II Testbed Project: a digital library service model* (Washington, D.C.: Council on Library and Information Resources, 1999), <http://www.clir.org/pubs/abstract/pub87abst.html>.

³⁴ Anne Gilliland-Swetland, *Enduring paradigm, new opportunities: the value of the archival perspective in the digital environment* (Washington, D.C.: Council on Library and Information Resources, 2000).

³⁵ *The London Charter for the computer-based visualisation of cultural heritage* (Londoncharterorg, 2009), <http://www.londoncharter.org>.

metaandmete skeeme. Metaandmete skeem on kindlal eesmärgil koostatud metaandmete elementide loetelu koos loogiliste seostega elementide vahel. Sellest tulenevalt on nii eri institutsioonid (raamatukogud, arhiivid, muuseumid, ettevõtted jne) kui ka neid koondavad huvigrupid loonud endale kõige sobivamad metaandmete skeemid ja standardid. Metaandmed on tavaliselt mingil viisil struktureeritud. Selleks määratletakse infoobjektide kirjeldamisel kasutatavad kategooriad (näiteks "Autor", "Pealkiri", "Teema"), info esitamise viis, kategooriate omavahelised seosed jms. Metaandmete struktuur vastab kirjeldatavate infoobjektide olulisematele tunnustele.

Kuna eri institutsioonid kirjeldavad erinevaid objekte, siis on ka kasutatavad metaandmete skeemid erinevad. Metaandmed võivad inforessursse kirjeldada erineval tasemel: kirjeldatakse nii kogusid tervikuna, üksikuid infoobjekte kui ka infoobjekti eri osasid, näiteks artiklit ja artiklis asuvat fotot.

Raamatukogudes kasutatakse objektide kirjeldamiseks bibliikirjeid, mille koostamisel lähtutakse kirje- ja kataloogimisreeglitest. Teavikud kirjeldatakse bibliograafilise kirjega ja kataloogitakse bibliograafilistes vormingutes. Kataloogimisreeglite aluseks on tänapäeval üldiselt ISBD (*International Standard Bibliographic Description*), mis on Rahvusvahelise Raamatukogude Ühenduste ja Institutsioonide Föderatsiooni (International Federation of Library Associations and Institutions, IFLA) välja töötatud standardite rühm, mis määrab eri teavikutüüpide kirjeldamisel esitatavate andmete valiku, vormi ja järjekorra, aga ka kirjavahemärgid, mida kirjes kasutatakse. Lisaks ISBD-le on olulised ka angloameerika reeglid (*Anglo-American Cataloguing Rules*, AACR). Nende teine, uuendatud väljaanne AACR2 põhineb kirje osas ISBD-l, kuid sisaldab lisaks pealdiste ja otsitunnuste moodustamise reegleid. AACR2 edasiarenduse tulemusena valmis 2010. aastal RDA (*Resource Description and Access*) standard. Tegemist on juhendite kogumiga, mis määratleb raamatukogudes, muuseumides jm säilitatavate teavikute kataloogimise reeglid.

Bibliograafilistest vormingutest on kõige enam levinud MARCi (*Machine Readable Cataloguing*) rühma kuuluvad standardid. MARCi standardid koosnevad MARCi vormingutest, mis määratlevad bibliograafilise teabe digitaalse esitamise ja vahetamise reeglid. Algne MARC loodi aastail 1965–66 Ameerika Ühendriikide Kongressi Raamatukogus. Eri riikides võeti kasutusele sellele tuginevad, kuid kohalikku kataloogimistava ja muid asjaolusid arvestavad n-ö riiklikud MARCi vormingud, näiteks Kanadas CAN/MARC, Saksamaal MAB jne. Tänapäeval kasutatakse ISO

standardit “ISO 2709:2008 Information and documentation – Format for information exchange”. Standardis on määratletud kirje põhistruktuur ja mitmesuguste tehniliste andmete esitusviis. Andmesisu määratlevad MARCi-välised reeglistikud ja standardid nagu ISBD ja AACR2. Kõige uuem vorming selles rühmas on MARC 21, mis tekkis kahe vormingu, USMARCi ja CAN/MARCi ühtlustamise tulemusena 1999. aastal.

Osades raamatukogudes, näiteks Tallinna Ülikooli Akadeemilises Raamatukogus ja Tartu Ülikooli Raamatukogus tuginetakse seal säilitatavate digitaalsete objektide kirjeldamisel Dublin Core'i metaandmestandardile. 1995. aastal loodi Dublin Core Metadata Initiative (DCMI), mille eesmärk oli elektroonilise info kirjeldamise standardi loomine. DCMI sai alguse 1995. aastal Iowa osariigis Dublinis toimunud konverentsil. 1998. aastal avaldati 15 elemendist koosnev Dublin Core Metadata Element Set (DCMES). “Informatsioon ja dokumentatsioon: Dublin Core'i metaandmeelemendid” on rahvusvaheline standard ISO 15836:2003, mis on üle võetud Eesti standardiks EVS-ISO 15836:2004. Dublin Core'i tunnustatakse ametlikult ka WWW Consortiumi ja ISO 23950 raames ning Dublin Core'i metaandmete elemendid on kehtestatud Ameerika Ühendriikide standardina ANSI/NISO Z39.85. Standard sisaldab metaandmete loetelu valdkondadevahelise elektroonilise informatsiooni kirjeldamiseks. Dublin Core'i 15 elementi jagunevad kolme üldisesse gruppi:

- inforessursi sisuga seotud elemendid;
- inforessursi kasutamise ja intellektuaalomandi õigustega seotud elemendid;
- inforessursi loomega seotud elemendid.

Dublin Core'i on võimalik tänu selle ülesehitusele laiendada, kombineerides sellesse teisi metaandmesüsteeme. Samuti saab mahukamaid süsteeme ekspordiks või süsteemidevaheliseks otsinguks Dublin Core'i teisendada. Dublin Core'i lihtsus on üheaegselt nii selle süsteemi tugev kui ka nõrk külg. Lihtsus vähendab metaandmete loomisega seotud kulutusi ja soodustab koostalitlusvõimet. Teisest küljest aga ei võimalda lihtsus sellist semantilist ega funktsionaalset mitmekülsust, mida toetavad keerukad metaandmesüsteemid. Samas saab keerukamaid metaandmesüsteeme vajadusel Dublin Core'iga kombineerida. Dublin Core ongi mõeldud kooskasutuseks teistsuguse semantikaga metaandmete standarditega, sageli ühes ja samas inforessursikirjelduses. Raamatukogudes kasutatavate

metaandmesüsteemide eripära seisneb selles, et kõik kogus leiduvad teavikud on vähemalt minimaalsel tasemel kirjeldatud.³⁶

Rahvusvaheline Arhiivinõukogu (ICA) on välja töötanud arhiivikirjelduse rahvusvahelise standardi ISAD(G), mille esimene versioon avaldati 1994. aastal. ISAD(G) kehtestab arhivaalide kirjeldustasandid ja määrab kirjelduselementide koosseisu vastavalt kirjeldustasandile. 2000. aastal avaldati selle teine, korrigeeritud väljaanne ISAD(G)2, mis on ka praegu kehtiv arhiivikirjelduse standard (ICA 2000). Standardis kasutatakse mitmetasandilist kirjeldussüsteemi, kus eristatakse järgmiseid tasemeid: arhiiv, kollektsioon, sari, säilik, arhivaal. Standard tugineb järgmistele põhimõtetele:

- kirjeldus liigub üldiselt üksikule;
- kirjeldusinfo peab vastama kirjeldustasandile;
- eri tasandite kirjeldused peavad olema seotud, st igal hetkel peab olema selge, millise kirjeldustasandiga on tegemist;
- kõrgemal kirjeldustasemel antud informatsiooni madalamal tasemel ei korrata.

Rahvusvaheline arhiivinõukogu on vastu võtnud ka standardi (ICA 2004), mis käsitleb normikirjete koostamist juriidiliste ja füüsiliste isikute ja perekondade kohta ning mille põhjal kirjeldatakse arhiivimoodustajaid. Lisaks on olemas ka rahvusvaheline standard ICA 2007, mis reguleerib juriidiliste isikute tegevuste ja funktsioonide kirjeldamist.

Kodeeritud arhiivikirjeldus (*Encoded Archival Description*, EAD) on metaandmete standard, mida kasutatakse peamiselt arhiivide, aga ka raamatukogude ja muuseumide kogude masinloetavate kirjelduste (inventari-raamatud, registrid, indeksid, kataloogid jms) loomiseks. EAD arendamine algas California Ülikoolis 1993. aastal. Standardi töötasid välja Ameerika Ühendriikide Kongressi Raamatukogu ja Ameerika Arhivaaride Ühing (Society of American Archivists). EAD tugineb SGMLile (*Standard Generalized Markup Language*) ja alates 1998. aasta versioonist EAD 1.0 on see kooskõlas ka XMLiga. EAD toetab igati arhiivide mitmetasandilist kirjeldussüsteemi ja on laiendatav <odd>-elementidega, kuhu saab mahutada sisu, mis teiste elementide alla ei mahu. EAD sisaldab 163 andmeelementi, mida saab kasutada info kodeerimiseks ja vahetamiseks. Paljud EAD-elementid on samased ISAD(G), MARCi ja Dublin Core'i standarditega. Eesti Rahvusringhäälingu arhiivis pikaajaliselt säilitatavad digitaalsed objektid

³⁶ Eluala Roel, "The MOSC project: using the OAI-PMH to bridge metadata cultural differences across museums, archives, and libraries", *Information Technology and Libraries*, 24:1 (2005), 22–24.

on kirjeldatud vastavalt Euroopa Ringhäälingute Liidu standarditele.³⁷ Arhiivid erinevad raamatukogudest selle poolest, et kõige paremini on tavaliselt kirjeldatud arhiivi-, sarja- ja säilikutasandid ning väga oluline on säilitada teabe igakülgne kontekst.

Ühtse struktureeritud metaandmestiku kasutuselevõtt Eesti muuseumides sai alguse 1992. aastal, kui Kultuuriministeeriumi tellimisel hakkas tarkvaraettevõtte AS GenNet Laboratories arendama KVISi (kultuuri-väärtuste infosüsteemi). KVISi väljatöötamisel võeti aluseks ICOMi (Rahvusvahelise Muuseuminõukogu) CIDOCi (International Committee for Documentation) andmemudel ning Rootsi atribuutide nime moodustamise standard SWETERM.³⁸ Tegemist oli objektorienteeritud andmemudeliga ning kirjelduse keskmes ei olnud objekt, vaid sündmus. Sellist kirjeldusmodelit toetab ka 2006. aastal CIDOCi poolt kasutusele võetud CRMi (*Conceptual Reference Model*) mudel. 2005. aastal võttis Kultuuriministeerium vastu otsuse luua uus muuseumide infosüsteem MuIS. Seejuures jäi uue infosüsteemi aluseks olev andmemudel samaks. Museaali kirjeldamist toetavad kesksed sõnastikud, mis tagavad museaali nõuetekohase kirje ja muuseumideülesed otsingud.³⁹ Loodusteaduslike kogude kirjeldamiseks kasutatakse lisaks ka muid infosüsteeme, näiteks maateaduste infosüsteemi SARV.⁴⁰ Eraldi andmebaasid spetsiifiliste metaandmetega on ka arheoloogilistel kogudel.⁴¹ Eesti Kunstimuuseumi digitaalkogus kasutatakse spetsiaalselt selle jaoks välja töötatud metaandmete süsteemi.

Seega on vaatamata kokkulangevustele erinevates mäluasutustes kasutusel eri metaandmete süsteemid ja standardid. Kuna eri institutsioonid on arendanud välja endale kõige sobivamad metaandmete süsteemid ja ühtset universaalset metaandmete süsteemi ei ole võimalik luua, siis on vajalik tagada metaandmete jagatavus eri süsteemide vahel. Tehnilisel tasemel sõltub koostalitlusvõime standardiseeritud süntaksite olemasolust ja ühiste kommunikatsiooniprotokollide kasutamisest. Standardsed süntaksid põhinevad tänapäeval XMLil ning levinumad kommunikatsiooniprotokollid on Z39.50 (ANSI/NISO Z39.50-2003) ja OAI-PMH (*Open Archives*

³⁷ EBU Technology & Innovation – Metadata Specifications, <https://tech.ebu.ch/MetadataEbuCore>.

³⁸ Kaie Jeaser, *Museaalide dokumenteerimine Eesti muuseumides*. Magistritöö info- ja kommunikatsiooniteaduse erialal (Tartu: Tartu Ülikool, 2009), 36–37, <http://hdl.handle.net/10062/14694>.

³⁹ *Museaali kirjeldamine MuISis*. Juhend 1.1. (Muinsuskaitseamet, 2019), https://www.muinsuskaitseamet.ee/sites/default/files/content-editors/Muuseumid/museaalide_kirjeldamisjuhend_2019.pdf.

⁴⁰ Eesti Geokogude Portaali: <http://geokogud.info>.

⁴¹ Eesti arheoloogia andmebaasid: <https://sisu.ut.ee/arheoloogia-andmebaasid/kogud>.

Initiative Protocol for Metadata Harvesting). Võrreldes tehnilise koostalitlusvõimega on märksa keerukam tagada semantiline koostalitlusvõime. Selleks on vajalik eri valdkondades kasutatava terminoloogia ja tähenduste ühtlustamine. Lihtsaim lahendus semantilise koostalitlusvõime tagamiseks on kasutada eri metaandmesüsteemide vastavustabeleid (*mappings* ehk *crosswalks*) kombineerituna valdkondadevaheliste metaandmestandarditega, näiteks Dublin Core'iga.

Koos mäluasutustes leiduva teabe digitaliseerimise ja elektrooniliste otsivahendite kasutuselevõtuga on oluliselt kasvanud metaandmete tähtsus kogu infosüsteemi toimimises. Lihtsalt väljendudes ei ole kasutajal enam kõrval kõiketeadjat raamatukogutöötajat, arhivaari või koguhoidjat, kes tunneb kogu nagu oma tagataskut. Kasutajate teadmatuse tõttu kipub osa teavet jääma n-ö varju ja kättesaamatuks. Selle vältimiseks tulebki ennekoike keskenduda teavikuid kirjeldavatele metaandmetele, kuna nende kaudu saadakse kätte vajaminev teave. Probleem peitub aga selles, et metaandmete loomise põhimõtted ja lähenemised erinevad valdkondade, institutsioonide ja isegi isikute vahel.⁴²

Eeltoodud metaandmete määratluste ja nende funktsioonide kirjelduste põhjal selgub, et metaandmete puhul on tegemist kindlat tüüpi informatsiooniga.⁴³ Laskumata siinkohal infomatsiooni termini määratlemise keerukusesse,⁴⁴ võib tõdeda, et info on alati seotud kommunikatsiooniga. Mõned uurijad defineerivadki infot kui midagi sellist, mis jääb kommunikatsiooniprotsessis muutumatuks.⁴⁵ Järgnevalt annan ülevaate kommunikatsiooniprotsessi käsitlemisel kasutatavatest mudelitest.

⁴² Vt Lawrence Dowler, "The Role of Use in Defining Archival Practice and Principles: A Research Agenda for the Availability and Use of Records", *American Archivist*, 51:1 (1988), 74–86; Jennifer Schaffner, *The Metadata is the Interface: Better Description for Better Discovery of Archives and Special Collections: Synthesized from User Studies* (Dublin: Online Computer Library Center, 2009), <http://www.oclc.org/programs/publications/reports/2009-06.pdf>.

⁴³ Luciano Floridi, *Information: a very short introduction* (Oxford: Oxford University Press, 2009), 31.

⁴⁴ Vt nt Giovanni Sommaruga, "One or many concepts of information?", *Formal theories of information: from Shannon to semantic information theory and general concepts of information*, ed. by Giovanni Sommaruga (Berlin, Heidelberg: Springer Verlag, 2009), 253–267.

⁴⁵ Einar Duenger Böhn, *God and abstract objects* (Cambridge: Cambridge University Press, 2019), 35.

Kommunikatsiooniprotsessi mudelid

Informatsioon ei esine ühiskonnas mingi eraldiseisva olemina, vaid on haaratud paljudesse protsessidesse, mis moodustavad ühiskondliku infosüsteemi. Selle infosüsteemi elementaarprotsessiks võib lugeda kommunikatsiooniakti. Kommunikatsiooni on defineeritud kui sündmuste järjestust, mis viib informatsiooni ülekandeni infoallikalt vastuvõtjale.⁴⁶ Informatsiooni ülekandmise tähistamiseks kasutatakse ka terminit “transmissioon”. Erialases terminoloogias on “kommunikatsioon” ja “transmissioon” sünonüümid. Sotsiaalses käsitluses eeldab kommunikatsioon teatest arusaamist. Signaalid võidakse küll edastada, aga kui vastuvõtja ei ole võimeline neid tõlgendama ja neile reageerima, siis kommunikatsiooni ei toimu. Siinkohal tasub tähele panna asjaolu, et vastuvõtja ei pea sõnumit tõlgendama samamoodi kui saatja, oluline on vaid see, et tõlgendamine toimuks.

Kommunikatsioon tugineb sümbolite transmissioonile ja nende sümbolite tõlgendamisele teate vastuvõtja poolt. Kuigi harilikult käsitletakse kommunikatsiooni inimeste vahel toimuva protsessina, ei tohiks unustada, et kommunikatsioonivõimelised on kõik elusolendid ja ka inimese loodud masinad. Kommunikatsioonisüsteemi kõige lihtsamal mudelil, mida nimetatakse stiimuli-reaktsiooni mudeliks (*SR model*, *stimulus-response model*), võib eristada teate saatjat, teate vastuvõtjat ja sõnumit (joonis 1).



Joonis 1. Kommunikatsioonisüsteemi lihtsaim, nn stiimuli-reaktsiooni mudel. Joonised kogu artiklis: Aide Eendra

Infoallikas ehk teate saatja võib olla nii inimene, mõni muu elusorganism kui ka masin, samuti võib infoallikana käsitleda inimeste gruppe ja institutsioone. Sõnum on infoallika poolt moodustatud tähendust kandev märkide kogum. Märkideks võivad olla häälikud, kirjatähed, visuaalsed kujundid vms. Sõnum võib olla suuline, kirjalik, pildiline vms. Sõnumi loomine võib iseenesest olla keeruline protsess ja selles võib olla mitmeid osalisi – mõeldagu siinkohal kasvõi raamatu või kinofilmide valmistamise peale. Vastuvõtja ehk adressaat võib olla inimene või mõni muu olem.

Täpsema analüüsi huvides tuleks eristada protsesse, mis kommunikatsiooniahelas toimuvad. Infoallikas koostab oma sisemistele seisunditele

⁴⁶ Mark Burgin, *Theory of Information: Fundamentality, Diversity and Unification* (New Jersey, London et al: World Scientific Publishing, 2010), 256.

tuginedes sõnumi, millel on sisu – tähendus infoallika ja adressaadi jaoks. Sõnum on alati esitatud mingil füüsilisel kujul, mille valib info looja. Seega valitakse info tekitamisel nii selle sisu kui ka vorm ehk esitusviis. Sisu saab edasi anda erinevas vormis. Tasub tähele panna, et sõnumi sisu ja vorm ei ole täiesti sõltumatud, vaid mõjutavad teineteist suurel määral.⁴⁷

Informatsiooni edastatakse, töödeldakse ja salvestatakse signaalide kujul. Signaalid on ajas muutuvad füüsilised suurused, nagu õhurõhk, elektripinge, võnkesagedus, valguse heledus vms. Signaalide abil on võimalik luua sidet ruumiliselt üksteisest kaugel asuvate kommunikatsiooniosaliste vahel. Signaale saab salvestada ja see võimaldab siduda ka ajaliselt eraldatud kommunikatsiooniosalisi. Infoallika loodud algse sõnumi transmissiooniks muudetakse see saatjas signaalideks; vastuvõtjas toimub jällegi vastupidine protsess: signaalidest saadakse sõnum. Signaalid võivad olla pidevad (nt analoogtelevision), diskreetsed (digi-TV) või kombineeritud (nt impulss-koodmodulatsiooni teisendused telefonivõrgus).

Vastavalt sellele, millisel füüsilisel kujul info esitatakse, on vajalik sobiv saatja. Kitsamas mõttes on info saatjaks looduslik või tehniline abivahend, mis teeb võimalikuks info esitamise kindlal viisil (joonis 2). Kui infot esitatakse suulise kõne kujul, võib saatjaks olla nii inimese hääleaparaat kui ka kõnesüntesaator. Adressaadil on vajalik jällegi vastuvõtja olemasolu, olgu selleks siis näiteks kõrv või mikrofoni.

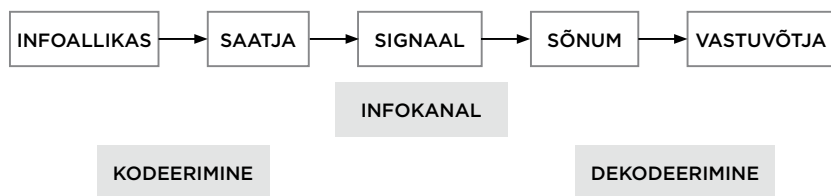


Joonis 2. Kommunikatsioonisüsteem, kus on eristatud info saatja

Sõnumi edastamine on ülekandeprotsess, milleks kasutatakse eri vahendeid: helilaineid, postisüsteemi, telegraafijuhtmeid, raadiolaineid, valgus-signaale vms. Seejuures võib info ülekanne toimuda nii ruumis kui ka ajas. Informatsioon on alati kodeeritud. Kodeerimine on informatsiooni esitusvormi muutmine kindla reeglistiku alusel. Reeglistikku, mis määrab kindlaks koodtähte moodustamise korra, nimetatakse koodiks. Koodid on kokkuleppelised ja ajalooliselt väljakujunenud. Näiteks morsetähestik on telegraafikood, milles tähtede, numbrite ja kirjavahemärkidega on vastavusse seatud pikkade ja lühikeste signaalide kombinatsioonid, näiteks

⁴⁷ Vt nt Kurmo Konsa, Edith Hermann, “The logic of text and structure of books: Changing the functions”, *Reading in changing society*, ed. by Marju Lauristin, Peeter Vihalemm (Tartu: University of Tartu Press, 2014), 28–38.

a-tähele vastab lühike ja pikk signaal (·, –). Kood koosneb märkidest ja reeglitest, mis määravad selle, kuidas ja mis seostes märke kasutatakse ning kuidas nendest kombineeritakse keerukamaid teateid. Kasutatava koodi alusel eristatakse suulist, kirjalikku ja visuaalset kommunikatsioonivormi. Erinevaid kodeerimis- ja dekodeerimisprotsesse võib kommunikatsiooniprotsessis olla mitu (joonis 3). Mida kauem infot säilitatakse, seda tõenäolisem on erinevate kodeerimis- ja dekodeerimisprotsesside esinemine selle elutsüklis.

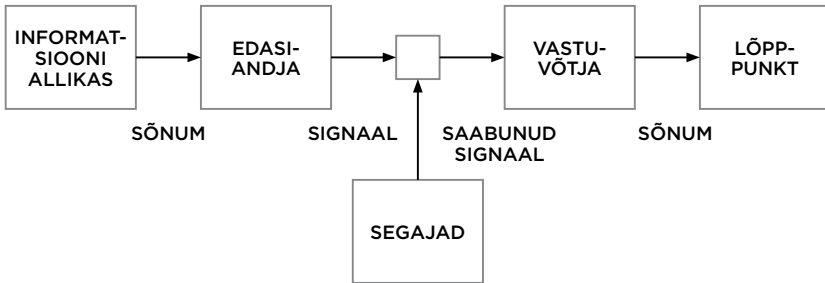


Joonis 3. Kodeerimis- ja dekodeerimisprotsesside eristamine kommunikatsiooniprotsessis

Kodeerimine ja dekodeerimine on iseenesest jällegi kommunikatsiooniaktid. Nagu edaspidi näeme, ei kulge kommunikatsioon vigadeta. Mida enam on kommunikatsiooniakte, seda suurem on vigade tõenäosus info ülekandel. Signaalid kanduvad edasi mööda side(info)kanalit. Kanalik on igasugune füüsikaline keskkond, milles signaalid levivad: valguslained, häälevõnked, raadiolained, närvirakud, valguskaablid jne. Sidekanali füüsikalised omadused määravad mingis ulatuses ära ka koodi, mida saab kasutada.

Signaalide ülekanne kanalis ei toimu tegelikult kunagi veatult, kuna alati kaasneb signaaliga ka müra. Müraks võib lugeda sidekanalis toimuvaid kõikvõimalikke protsesse, mis moonutavad või häirivad edastatavaid signaale. Müra saab jagada väliseks müraks, mille korral asub müraallikas vastuvõtjast väljaspool, ja sisemiseks müraks, mille korral on müraallikas vastuvõtja ise. Väliseks müraks võivad olla näiteks elektromagnetilised häiritused või liiga vähene valgustatus. Sisemine müra on näiteks elektronseadmetes tekkiv soojuslik müra, aga ka psühholoogilistest ja kultuurilistest teguritest tingitud sõnumite vastuvõtuhäired. Sõnumi sumbumise põhjuseks võib olla ka selle mittevastavus kasutatava sidekanali omadustele. Selles olukorras võib tekkida näiteks siis, kui sõnumis sisalduv informatsiooni-hulk ületab kanali läbilaskevõime. Kirjeldatud kommunikatsioonimudeli,

mida tuntakse ka edasikandelise (*transmissional*) mudeli nime all (joonis 4), esitas Claude E. Shannon oma 1948. aastal ilmunud artiklis “Kommunikatsiooni matemaatiline teooria”.⁴⁸



Joonis 4. Shannoni kommunikatsioonimudel ehk info transmissioonimudel

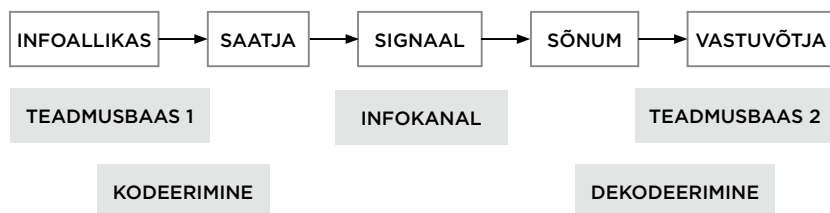
Kommunikatsioonikanali tähenduses kasutatakse sageli ka terminit “meedium”. Meediumi all mõistetakse nii kommunikatsioonikanalit, ülekandjat ennast kui ka vastavaid organisatsioone. Nii on meediumiteks näiteks kõne, raamat, film, internet, aga ka kirjastused ning raadio- ja televisiooni-organisatsioonid. Shannoni mudel ei tegele otseselt sõnumite tähendusega. Ometigi vahetame me sõnumeid ainult sellepärast, et need tähendavad midagi. Sõnumi tähendus tekib sõnumi ja selle tajuja vastastikmõjus. Selleks, et informatsioon saaks tähenduse, peab seda kasutama kas inimene, mõni muu elusolend või inimese loodud tehisklik süsteem.

Tähendust esitab atomaarne kommunikatsiooniakt tervikuna toimides: teave ei esine selle süsteemi üheski osas eraldi, vaid kogu süsteemis, sealhulgas selle süsteemi toimimises, st protsessis endas. Tähendus on lahutamatu seotud kontekstiga, milles kommunikatsioon toimub. Info loomine ja sellest arusaamine nõuab kontekstuaalset teavet. Kontekst tähendab seejuures nii konkreetse kommunikatsiooniolukorra keskkonda kui ka kindlale sõnumile eelnenud ja järgnenud teateid ning nende tähendusi. Märgid ise ei kannu tähendust isoleerituna, vaid kindla koodisüsteemi liikmena. Laiemas käsitluses hõlmab kontekst tervet maailma, kus kommunikatsiooniakt aset leiab. Kontekst mõjutab kogu kommunikatsiooniakti.

⁴⁸ Claude E. Shannon, “A mathematical theory of communication”, *The Bell System Technical Journal*, 27 (1948), 379–423, 623–656, <http://math.harvard.edu/~ctm/home/text/others/shannon/entropy/entropy.pdf>.

Konteksti olulisust sõnumi tähenduse mõistmiseks rõhutas oma kommunikatsioonimudelil semiootik Roman Jakobson, kes lisas kommunikatsioonimudelisse ka kontakti mõiste.⁴⁹ Selleks, et kommunikatsioon saaks üldse toimuda, peab olema kontakt saatja ja vastuvõtja vahel. Kontakt tähendab siinkohal nii füüsilist kommunikatsioonikanalit kui ka psühholoogilist kontakti sõnumi saatja ja vastuvõtja vahel.

Saatja ja vastuvõtja peavad kommunikatsiooniolukorda sarnaselt tõlgendama. Seega on kommunikatsiooni toimumiseks vajalik, et sõnumi saatjal ja vastuvõtjal oleksid ühesugused teadmised selle kohta, kuidas sõnumid on kodeeritud ja kuidas nad tähendustega seostuvad. Loomuliku keele abil suhtlemiseks, näiteks vestlemiseks, on vajalik osata ühist keelt. Ühine keel tähendab aga lisaks ühisele koodile ka suure hulga ühiste teadmiste olemasolu. Infoallikal ja adressaadil on kindlad teadmised, millele tuginedes sõnumid luuakse ja neist aru saadakse. Sellist teadmiste kogumit nimetatakse teadmusbaasiks (*knowledge base*) (vt joonis 5).



Joonis 5. Infoallika ja adressaadi teadmusbaasid info kommunikatsioonisüsteemis

Raamatust arusaamine eeldab keele- ja kirjaoskust, selle sisu täpsem mõistmine võib nõuda teadmisi ka konkreetsest teemavaldkonnast (ajaloost, teoloogiast, füüsikast jne). Minevikust või teistest kultuuridest pärit tekstide korral on vajalik lisaks veel ajaloolise ja kultuurikonteksti tundmine. Teadmised muutuvad loomulikult koos ühiskonna ja kultuuri arenguga. Kui infoallika ja adressaadi teadmusbaasid on täiesti erinevad, ei saa kommunikatsioon toimuda. Kui aga teadmusbaasid täielikult kokku langevad, pole kommunikatsioonil erilist mõtet, kuna informatsiooni hulk, mida kommunikatsioon suudaks pakkuda, oleks minimaalne. Tõsi küll, mõnikord

⁴⁹ Roman Jakobson, "Linguistics and poetics", *Selected Writings III. Poetry of Grammar and Grammar of Poetry* (The Hague: Mouton Publishers, 1981), 18–51 (22, 27).

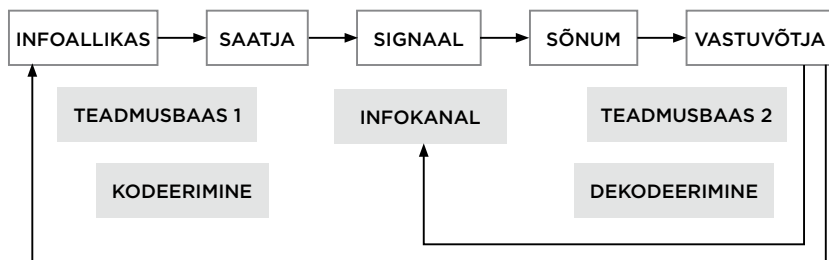
võib vähene informatsioon olla suure kaaluga, näiteks kui saadetud sõnum on mingiks oluliseks tegevuseks vajalik "võti". Saadetud sõnum võib vastuvõtja teadmusbaasis mingi muudatuse esile kutsuda ja kui see muudetud info on tihedalt seotud ülejäänud teadmusbaasiga, siis võib vastuvõtja teadmusbaas ahelreaktsioonina tähelepanuväärselt teiseneda. Piisavalt erinevat teadmusbaasi nõudva info mõistmiseks kasutatakse seda valdavate spetsialistide abi. Näiteks vanakreeka keeles kirjutatud tekste vahendavad, tõlgivad ja kommenteerivad vastava hariduse saanud spetsialistid – kõiki-dele teistele jäävad need mõistetamatuks.

Konteksti (tähenduste ja väärtuste) muutumine muudab info mõistmise võimatuks või vähemalt raskendab seda tunduvalt. Info loomise ja kasutamise kontekstid ei ole kunagi täiesti identsed. Mida kultuuriliselt erinevamad on info looja ja vastuvõtja, seda erinevamad on ka kontekstid. Kontekstide erinevused muutuvad eriti oluliseks asünkroonse kommunikatsiooni korral, kus sõnumi saatmise ja vastuvõtu vahele jääb ajaline distants. Mida suuremad on kontekstide erinevused, seda enam täiendavaid kommunikatsiooniprotsesse on kommunikatsiooni haaratud.

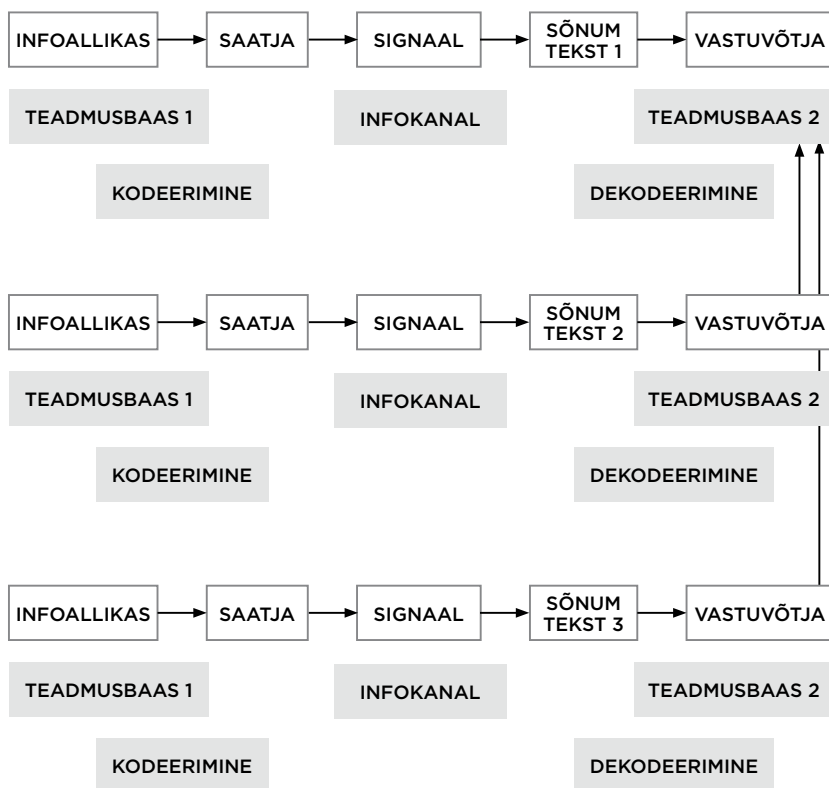
Shannoni kommunikatsioonimudel on ühesuunaline ja lineaarne. Tegelik kommunikatsioon sarnaneb aga sageli hoopis dialoogile, kus sõnumeid vahetatakse vaheldumisi ning sõnumite sisu ja vormi varieeritakse vastavalt kaasvestlejalt saadud sõnumitele. Sellise olukorra kajastamiseks lisatakse kommunikatsiooniprotsessi mudelisse tagasiside. Tagasiside on protsess, mille käigus sõnumi saatja saab infot selle kohta, kas vastuvõtja on sõnumi kätte saanud ja kuidas ta on seda mõistnud – kas sõnumil on olnud selline mõju nagu saatja eeldas. Otsese suhtluse käigus saame partnerilt pidevalt tagasisidet sõnade, žestide, miimika jms kaudu. Tagasiside võib olla ka keerukam, selleks võivad olla loodud spetsiaalsed institutsioonid – teaduskirjanduse eelretsenseerimine (*peer review*), kriitika, filmide eellinas-tused jms. Tagasiside võimaldab saatjal edastatavaid sõnumeid vastavalt adressaadi reageeringule kohandada. Kommunikatsiooniprotsessi tagasi-sideahelaid on oma mudelis lähemalt käsitlenud Brian ja Alina Vickery.⁵⁰

Saadetud ja vastuvõetud sõnumite erinevuse vähendamiseks võidakse muuta mingeid sidekanali omadusi, näiteks rääkida kõvemini või suurendada kanali läbilaskevõimet (vt joonis 6). Teise tagasisidestuse kaudu püütakse mõjutada saadetavat sõnumit ennast, näiteks palutakse sõnumi saatjat rääkida selgemini või öelda nimi tähthaaval. Tagasiside võib olla ka kaudsem ja toimuda hiljem, seda eriti asünkroonse kommunikatsiooni korral. Õpetaja

⁵⁰ Brian Vickery, Alina Vickery, *Information science in theory and practice* (London: Bowker-Saur, 1992), 41–42.



Joonis 6. Üks võimalik tagasisideahel kommunikatsiooniprotsessis



Joonis 7. Kommunikatsiooniahelate seostunud võrgustik

võib muuta oma sõnumit alles pärast vastuvõtu kontrollimist (tunnikontrolli), reklaamifirma muudab reklaami pärast ostutulemuste selgumist jne.

Elementaarsed kommunikatsiooniaktid grupeeruvad enamikul juhtudest seostunud kommunikatsiooniahelateks. Toome näiteks võõrkeelse teadusliku teksti lugemise (joonis 7; 1. ahel). Teksti lugedes kasutatakse sõnaraamatut (joonis 7; 2. ahel), lisaks kontrollitakse internetiotsingu abil terminite määratlusi (joonis 7; 3. ahel).

Teine ja kolmas kommunikatsiooniprotsess aitavad kasutajal teadusbaasi täiendamise teel esialgu arusaamatust tekstist aru saada. Kommunikatsiooniaktis muutub teate vastuvõtja seisund. Sageli moodustab vastuvõtja esimese kommunikatsiooniakti järel uue sõnumi või täiendab mingil viisil algset sõnumit.

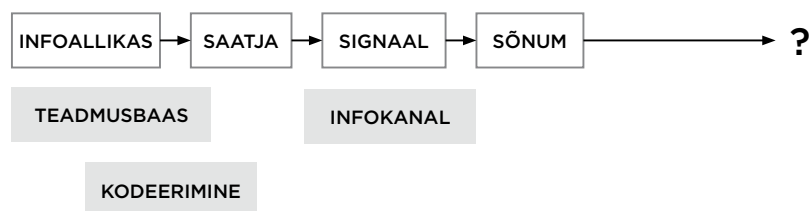
Kommunikatsioonivõrgustike toimimise tagavad kindlad institutsioonid. Need institutsioonid on kultuurilised mehhanismid, mille abil luuakse ja hoitakse toimimas tähenduslik ning korrastatud sotsiaalne maailm.⁵¹

Metaandmed kommunikatsiooniprotsessis

Pöördugem nüüd uuesti artikli põhiküsimuse juurde – millise kommunikatsiooniprotsessi osaga on siis seotud metaandmed. Vaatleme järgmist, esmapilgul lihtsat ja igapäevast olukorda. Mari ja Ants vestlevad mobiiltelefonis. Selle kommunikatsioonisituatsiooni kirjeldamiseks oleme me juba kasutanud mitmeid metaandmeid. Info allikaid ja vastuvõtjad on kirjeldatud nimedega. Meediumiks on suuline kõne ja signaal edastatakse raadiolainetena. Edasiandjaks ja vastuvõtjaks on tehnilised seadmed – mobiiltelefonid. Kui me soovime seda kommunikatsioonisituatsiooni täpsemalt kirjeldada, siis võime lisada täiendavalt metaandmeid. Isikuid saab täpsemalt kirjeldada, lisades perekonnanimes, isikukoodid, sünniajad ja -kohad jne. Toimumispaigad saab fikseerida geograafiliste koordinaatide kaudu, samuti saab fikseerida vestluse aja. Tehniliste metaandmetena saab lisada mobiiltelefonide margid, telefonis kasutatava opsüsteemi ja muud programmid, teenuseosutaja ja -paketi, telefoninumbrid jms.

Nagu näha on metaandmed seotud kommunikatsiooniprotsessi kõikide elementidega. Metaandmed on tegelikult neid elemente kirjeldavad tunnused (atribuudid). Iga kommunikatsiooniakti elemente iseloomustavad kindlad atribuudid, mille fikseerimine annab tervikliku teabe selle

⁵¹ David M. Levy, "Documents and libraries: a sociotechnical perspective", *Digital library use: social practice in design and evaluation*, ed. by A. B. Bishop, N. A. Van House, B. Pfeil Buttenfield (Cambridge, London: The MIT Press, 2003), 26–42, 38.



Joonis 8. Asünkroonne kommunikatsiooniprotsess

kommunikatsiooniakti kohta. Kommunikatsiooniakti elementide kõik atribuudid moodustavad selle kommunikatsiooniakti täieliku metakirjelduse. Tegelikult võib kommunikatsiooniakti elemente iseloomustada väga paljude tunnustega. Eeltoodud näites võib ju osalevaid inimesi iseloomustada ka füüsiliste parameetrite, geneetilise koodi jms kaudu. See, millised metaandmed on tähtsad, sõltub otseselt sellest, milleks neid kasutatakse. Telefonivestluses osalejal on üldjuhul vaja teada vaid selle inimese nime, kes talle helistab, samal ajal kui näiteks korrakaitseorganeid võivad huvitada ka kõnede toimumisajad ja -kohad.

Metaandmete seisukohalt on oluline võtta arvesse kommunikatsiooni ajalisi aspekte. Kommunikatsiooni võib jagada sünkroonseks ja asünkroonseks. Sünkroonse kommunikatsiooni korral on kõik osapooled korraga haaratud reaajas toimuvasse suhtlusprotsessi. Osalejad võivad ruumiliselt asuda samas kohas (näiteks otsene suuline vestlus kahe inimese vahel) või olla üksteisest füüsiliselt eraldi (näiteks telefonivestlus, videokõne, tekstisuhtlus, jututuba vms). Asünkroonne kommunikatsioon on ajalise viivitusega toimuv suhtlus, informatsiooni edastamiseks kasutatakse mingit vahendajat (kiri, kujutised jms). Asünkroonse kommunikatsiooni korral on osalised tavaliselt ka ruumiliselt eraldatud.

Asünkroonse kommunikatsiooni korral jääb joonisel näha olev kommunikatsiooniprotsessi vastuvõtja poolne osa (parempoolne küsimärgiga osa) võimaluseks (joonis 8). Kommunikatsiooniakti saatjapoolne sõnum on loodud ja me eeldame, et sõnumi vastuvõtt ja interpreteerimine toimub kunagi hiljem või kusagil mujal. Kuigi see võib jääda ka toimumata, eeldame me alati, et see ikkagi toimub.

Kui ma kirjutan praegu seda artiklit, siis toimub kommunikatsiooniprotsess minu ja arvuti vahel. Ülevaade selle kommunikatsiooniprotsessi elementide olulisematest tunnustest on toodud tabelis 1.

Tabel 1. Artikli kirjutamise kommunikatsiooniprotsessil osalevate elementide tunnused

Saatja	Sõnum	Infokanal/vastuvõtja	Kontekst
Nimi	Pealkiri	Faili nimi, vorming, failivormingu versioon, faili laiend, faili MIME tüüp, faili asukoht, faili maht	Loomise eesmärk
Isikukood, sünniaeg	Keel	Faili kaitse, krüpteeringud, faili püsivus, päise esimesed 50 baiti (<i>Raw header</i>)	Finantseerija(d), tellija(d)
Ametikoht	Teema	Font, mall, lehekülgede arv, lõikude arv, sõnade arv, märkide arv (tühikutega, ilma)	Kirjutamisel kasutatud kirjan-dus, muud lisa-materjalid
Haridus	Märksõnad	Faili päise osa, kus kirjeldatakse dokumendi osade grupe (<i>Headingpairs</i>)	Täiendavad märkmed
CV/ETISe andmed	Kokkuvõte	Kasutatud tekstiredaktor, selle versioon	
Loomise koht või kohad, riik, maakond, aadress, GPSi koordinaadid	Versioonid	Kodeering (failis kasutusel olev kodeering või märgikooditabel)	
Loomise aeg	Muudatused, nende tegemise ajad	Salvestuste ja printimiste ajad	
		Varukoopiad	
		Arvuti, opsüsteem, klaviatuur, hiir, protsessor, emaplaat, videokaart, monitor jms	

Osa tabelis toodud tunnustest salvestatakse automaatselt arvuti poolt (neid saab vaadata faili atribuutidest) ja osa lisan mina autorina. Faile ja kogu arvutiga seotud tehnilist süsteemi saab kirjeldada automaatselt. Küll on ras-kem luua sõnumit ja selle konteksti iseloomustavaid tunnuseid. Huvitavat

võimalust kirjutamisprotsessi psühhofüsioloogiliste metaandmete säilitamiseks pakuvad klahvivajutuste salvestamise programmid.⁵²

Milline osa nendest kommunikatsiooniprotsessi iseloomustavatest tunnustest valitakse välja metaandmeteks, sõltub otseselt metaandmete funktsioonist. Kui nüüd artikkel avaldatakse, siis kõige enam muutuvad infokanali ja vastuvõtjaga seotud tunnused. Kogu teave, mis on seotud selle dokumendi ja failidega, jääb artikli autori valdusesse. Osa saatjat, sõnumit ja konteksti iseloomustavat teavet aga lisatakse artiklile, samuti erinevatesse andmebaasidesse.

Artikkel on ette nähtud osalema laias kommunikatsiooniprotsessis, mis toimub autori ja lugejate vahel. Selle korral on aga tegemist selgelt asünkroonse kommunikatsiooniga. Hetkel, kui ma kirjutan artiklit, ei tea ma ju täpselt, kes on selle lugejad. Mul on olemas mingi eelteave ja ootused, mis väljenduvad kasvõi ajakirjas, kuhu ma kirjutan, nõuetes artiklile jms. Ometigi on mul võimatu ennustada tulevikus toimuva kommunikatsiooniprotsessi üksikasju. Just see asjaolu keerustab oluliselt teabe säilitamist tuleviku tarbeks. Enamiku kommunikatsiooniprotsesside korral ei ole nende ajaline ulatuvus oluline. Suurem osa teabest on harilikult määratud lühiajaliseks kasutamiseks. Tegemist võib olla info looja (indiviid, organisatsioon, masin) jaoks mingiks ajaks vajaliku teabega. Vaid osa teabest on aga selline, mida soovitakse säilitada võimalikult pikka aega. Selleks on välja kujunenud kindlad institutsioonid ja nendega seotud protsessid. Enamikku ühiskonnas käigus olevast teabest kasutatakse üha uuesti, töödeldakse ümber. Vahel harva on vajalik pöörduda tagasi info algallika juurde. Igapäevases elus me üldjuhul ei vaja paari aasta tagust e-kirjavahetust. Loomulikult sõltub kõik tegevusvaldkonnast, näiteks ajaloolased eelistavad alati tugineda algallikatele. Kui aga õpime koolis Archimedese seadust, siis me ei loe Archimedese originaalteoseid, seda enam, et neid ei olegi säilinud. Enamikul juhtudest ei takista info algallika kadumine selle edasist kasutamist, kuigi kaotus ajaloole ja kultuuriloole on loomulikult suur ning sageli korvamatu. Originaalallikate olemasolu on vägagi oluline aga näiteks selleks, et vajadusel oleks võimalik info autentsust kontrollida.

Selleks et säilitatav teave oleks kasutatav ja mõistetav peab sellega kaasnema teatud hulk metaandmeid. Koos digiinfo kasutuselevõtuga on muutunud oluliseks ka selle säilitamine. Just digiinfo korral on aga metaandmed säilitamise üheks sõlmküsimuseks. Selle põhjuseks on asjaolu, et digiinfo kasutatavus ja mõistetavus sõltub suurest hulgast täiendavast

⁵² Vt nt Olga Gaitšenja, *Venekeelsete eesti keele õppijate kirjutamisprotsess* (Tallinn: Tallinna Ülikool, 2019).

teabest kommunikatsiooniprotsessi tehniliste ja kontekstuaalsete üksik-
asjade kohta.⁵³

Kokkuvõte

Digitaalne kommunikatsioon on meie kaasaegse ühiskondliku infosüsteemi aluseks. E-post ja muud internetiteenused, mobiilsed infosüsteemid, raadio ja televisioon, filmid ning kujutised – kõik need põhinevad digitehnoloogial. Suur osa digitehnoloogiast jääb tavakasutajale tabamatuks, sest on muutunud nähtamatuks, taandudes tehnilistesse süsteemidesse. Suur osa tehnilistest seadmetest on tänapäeval varustatud mikroprotsessoritega, mis töötavad digitaalsel põhimõttel. Lisaks inimestevahelisele suhtlusele on tekkinud inimeste ja masinate ning masinate omavaheline kommunikatsioon.

Digitaalsed meetodid on andmete kogumise, töötlemise, kasutamise ja kommunikatsiooni muutnud enneolematult kiireks ja kõikehõlmavaks. Andmeteks on muudetud kõik ja need on sageli reaalsajas kasutatavad. Informatsiooni olulisuse kasvuga on järjest enam suurenenud ka metaandmete roll. Metaandmed kirjeldavad andmeid, loovad nendele konteksti ja on sellistena hädavajalikud andmete tõlgendamiseks ja kasutamiseks. Metaandmed on seejuures vajalikud nii inimestele kui ka arvutisüsteemidele. Metaandmed lisavad andmetele täiendava väärtuse ja mõnikord muudavad need üldse kasutatavateks. Kui digiinfoelsetes mäluasutustes täitsid metaandmed peamiselt kasutajat abistavat osa, siis digiteave on ilma nendeta väärtusetu ja kasutuskõlbmatu. Kui kasutada artiklis arendatud ideed metaandmete seosest kommunikatsiooniprotsessiga, siis võib väita, et isegi siis, kui meil pole enam andmeid, aga on säilinud metaandmed neid loonud kommunikatsiooniprotsessi kohta, on metaandmete abil võimalik vähemalt osaliselt see kommunikatsiooniprotsess taastada. Nii et metaandmed võivad olla kasulikud ka siis, kui andmed ise on kadunud. Selles artiklis esitasin idee, et metaandmed tulenevad kommunikatsiooni-
protsesside elementide tunnustest. Selline lähenemine võimaldab luua süsteemaatilisema aluse metaandmete valikuks ning täiendada automaatseid metaandmete loomise süsteeme.

⁵³ Vt nt Kurmo Konsa, *Bitid purki: teabe säilitamine digiühiskonnas* (Tartu: Kõrgem Kunstikool Pallas, 2018).

ABSTRACT: *The origin of metadata: a perspective on information theory*

According to the simplest and most common definition, metadata refers to a set of data that describes other data. Research on metadata is focused almost exclusively on solving practical issues. There are few theories on metadata that emphasise the lack of a common theoretical foundation to handle metadata, and there is also a lack of corresponding research. This article looks at metadata from a broad perspective of information technology and seeks an answer to a question that may, at first glance, seem simple: what is the origin of metadata? The article aims to present a conceptual model that connects metadata to communication processes, thereby creating an opportunity to treat metadata in a more systematic manner.

In memory institutions, different metadata schemes and standards are used to describe digital objects. In order to describe objects, libraries use bibliographical entries that correspond to valid entry and cataloguing rules. Objects are described by bibliographic entries and catalogued in a bibliographic format. Nowadays, cataloguing rules are mostly based on the ISBD (International Standard Bibliographic Description). The most common bibliographic formats are standards belonging to the MARC (Machine Readable Cataloguing) group. Some libraries, such as the Academic Library of Tallinn University and the University of Tartu Library use the Dublin Core metadata standard to describe the digital objects they preserve. A particular feature of the metadata systems used by libraries is that all objects in a collection are described to at least a minimum level. Archives in Estonia use the General International Standard Archival Description (ISAD(G)), which was developed by the International Council on Archives. Archives differ from libraries in that archives usually describe objects in detail on the levels of archive, series, and archive item, and preservation of the full context of information is prioritised. Estonian museums began introducing common structured metadata in 1992, when the Ministry of Culture commissioned a software company called AS GenNet Laboratories to develop KVIS (Information System of Cultural Values). The development of KVIS was based on the CIDOC (International Committee for Documentation) data model of the International Council of Museums, and on SWETERM, the Swedish standard of forming name attributes. This was an object-oriented data model, and the description was focused not on the object but the event. This type of description model is also supported by the CRM (Conceptual Reference Model) adopted by CIDOC in 2006. In

2005, the Ministry of Culture decided to create MuIS, a new information system for museums, although this new system was based on the same underlying data model as the previous system. The descriptions of museum objects are supported by central glossaries that ensure museum items are described as required and that searches can be made across museums. To describe natural scientific collections, other information systems are also used, such as the SARV database, used for managing data related to geo-collections. Archaeological collections have their own databases as well, with specific metadata. The digital collection of the Art Museum of Estonia uses a bespoke system of metadata.

Metadata are connected to each of the elements in the communication process: metadata are the attributes describing these elements. Each element of a communication act is characterised by specific, fixed attributes that provide full information about the act. All the attributes of the set of elements pertaining to a specific communication act make up the full meta description of this communication act. In fact, a communication act can be characterised by various attributes. The selection of metadata attributes used to characterise a communication process is connected to the function of the metadata relevant for the particular case.

KEYWORDS: metadata, descriptive metadata, memory institutions, communication process

KURMO KONSA is Associate Professor of Archival Studies at the Institute of History and Archaeology, University of Tartu.*

* Correspondence: Institute of History and Archaeology, University of Tartu, Jakobi 2, 51004, Tartu, Estonia. E-mail: kurmo.konsa@ut.ee

