

-
- # FormatE
-

- Nuläge och framtid
-

- Benjamin Yousefi

- *FAI (Föreningen för arkiv- och informationsförvaltning) oktober 2022 10:00 – 10:40*



Riksarkivet

Vem är jag?

Juridisk och teknisk rådgivare
och utredare på avdelningen
för informationshantering (AFI)



Benjamin.Yousefi@Riksarkivet.SE

- Utredningen för betrodda tjänster (2020-2021)
- Samrådsgruppen (2016-2021)
- FormatE (2018++)
- ArkivE (2013-2018)
- Elektroniska underskrifter (2013++)
- Preforma (2013-2018)
- PDF/A (2013++)

-
- Översikt
-
-
-
-
-
-

Vad ska jag prata om idag?

Översikt – Vad ska jag prata om idag?

Om

- 1. Vad händer?**
 - *Pågående arbeten och kommande steg*
- 2. Vad är FormatE?**
 - *En översikt av vad FormatE omfattar*
- 3. Vad är problemet?**
 - *Exempel på vad vi försöker uppnå med FormatE*

-
- 1. Vad händer?
-
-
-
-
-
-

Pågående arbeten och kommande steg

1. Vad händer?

- Vi är nu på år 1 av 5
- Webbsida **[HTTP://Riksarkivet.SE/FormatE](http://Riksarkivet.SE/FormatE)**

1. Vad händer? – Remissbehandlingen

- Intern diskussion om inkomna remissvar
- Arbetar med en skriftlig sammanställning och formella svar
 - Tar tid
 - Resultatet tas dock med i pågående arbeten. Till exempel:
 - Vägledningar
 - Färdigställa författningsförslagen
 - Sammanställningen tillsammans med ursprungliga remissvar kommer att publiceras på webbsidan

1. Vad händer? – Kommande steg

Remissbehandlingen knyter ihop ”den teoretiska säcken” och fokus är nu på:

- Främjande – kurser, seminarier, utbildningar, vägledningar
- Praktisk hjälp – program, referensimplementeringar och tekniska kontroller, handledningar
- Empiri – enkäter, fallstudier, specifikationer
- Samarbeten – specifika verksamheter om specifika frågor

1. Vad händer? – Arbetsgrupper och vägledning

- Elektroniska fakturor – Utkast till manus för publicering
 - Samarbete med SFTI
- Om beständighet och arkivbeständighet – ett första utkast färdigt – ”internremiss”
- Rörlig bild och video – Utkast under arbete
 - Kontakta intressenter nu i höst, vinter
- Fler på väg – informationsformat, kontrollfunktion, ...

-
- 2. Vad är FormatE?
-
-
-
-
-
-

En översikt vad FormatE omfattar

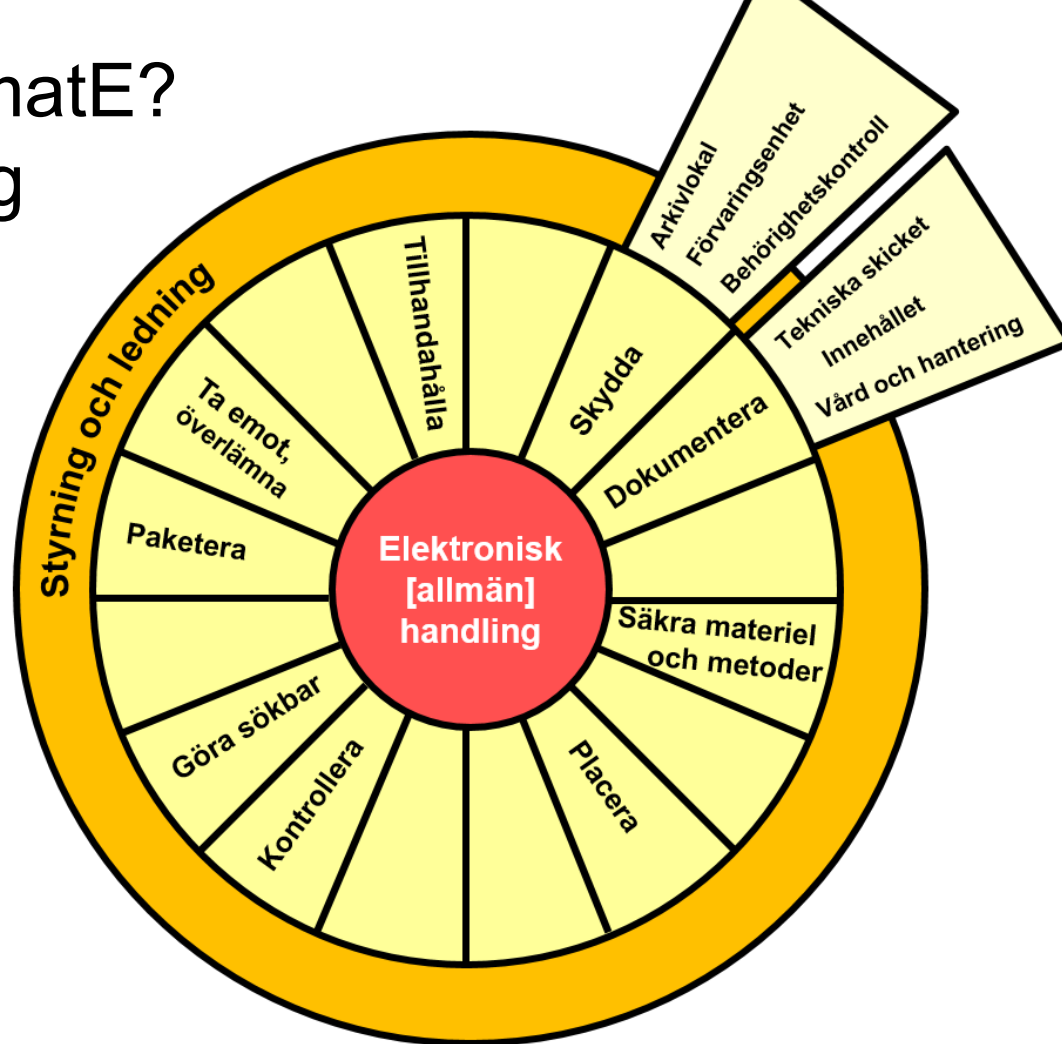
2. Vad är FormatE? – En ”nатурlag”

- FormatE är inget ”revolutionerande”, ingen ”ny idé”; varken en ”uppfinning” eller ”upptäckt”.
- Det är en *kartläggning av*, försök att *organisera, strukturera, systematisera vad som krävs av materiel och metoder för att koda, avkoda format konsekvent (beständighet)*.
- Det vill säga, en *metodlära*.

2. Vad är FormatE? – Avgränsning



2. Vad är FormatE? – Avgränsning



- "ArkivE"
- "Arkivbeständighet"

2. Vad är FormatE? – Avgränsning

- FormatE – Endast den *elektroniska handlingen*



2. Vad är FormatE? – Avgränsning

- FormatE – De fysisk-logiska (mätbara) ”delarna” som *representerar handlingen – det tekniska skicket*



- Beständighet – att kunna återge det tekniska skicket

2. Vad är FormatE? – Avgränsning

- Tekniskt skick – Grovt förenklat
 - Program + [ett eller flera] format
 - Varför? – Format är beroende av program
 - Inte alltid enkelt att skilja mellan program och format. Till exempel, CDF, databaser, webbsidor.

2. Vad är FormatE? – Avgränsning

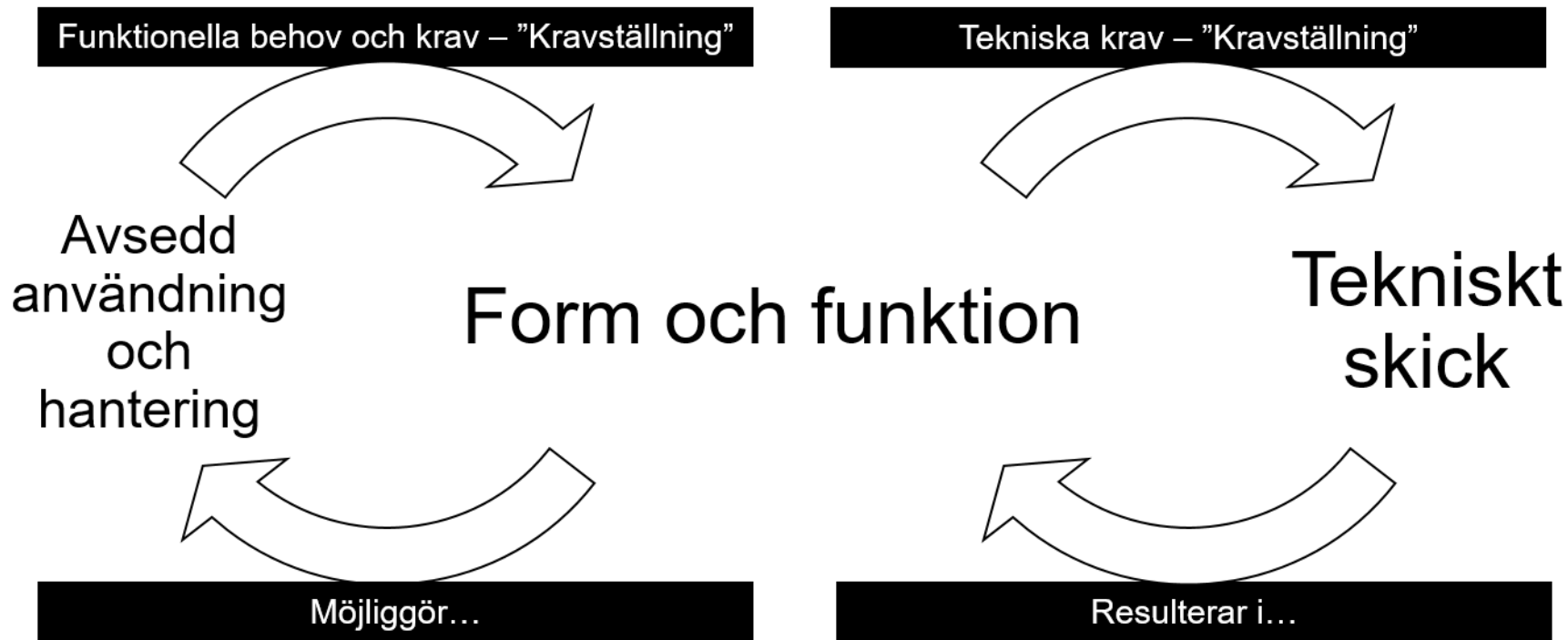
- FormatE omfattar inte heller egentligen det funktionella skicket av handlingen. *Det vill säga, innehållet*
- *Försöker dock identifiera gränsen mellan tekniskt skick och funktionellt skick*



Innehållet

- Fråga om värdering och gallring
- Vad som krävs för att uppfylla 3 § arkivlagen

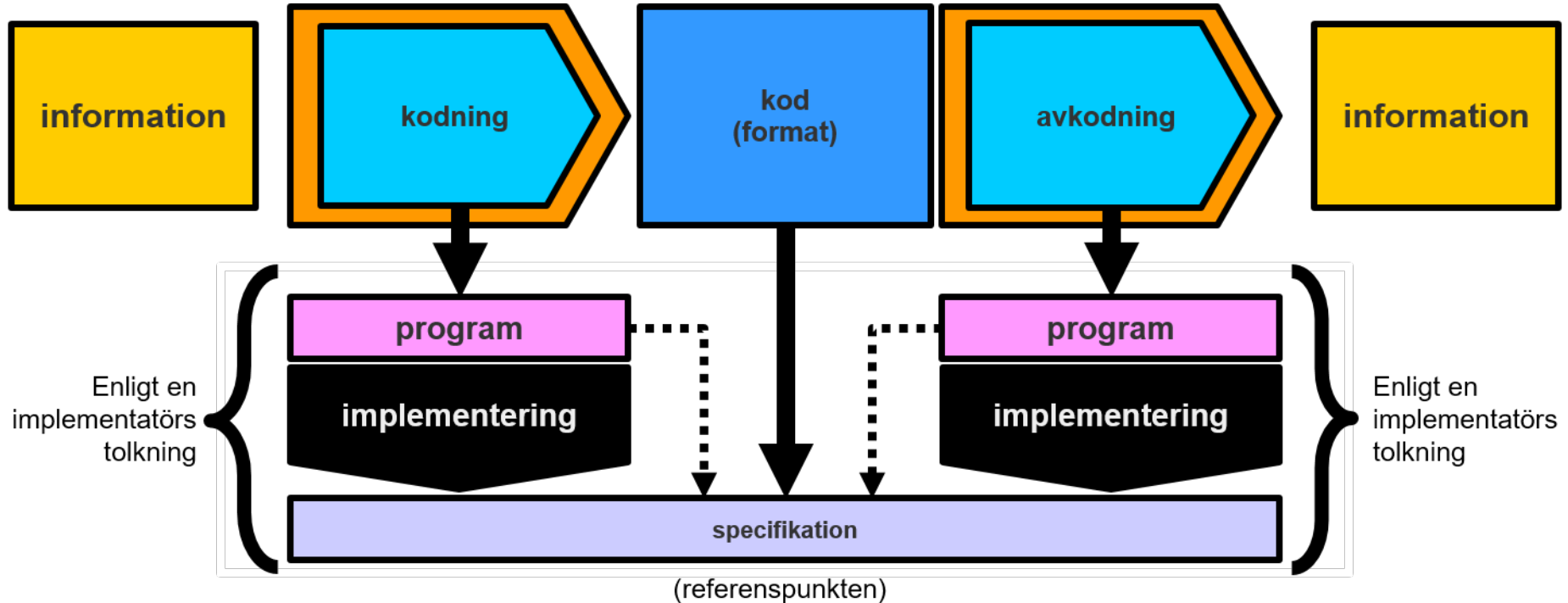
2. Vad är FormatE? – Avgränsning



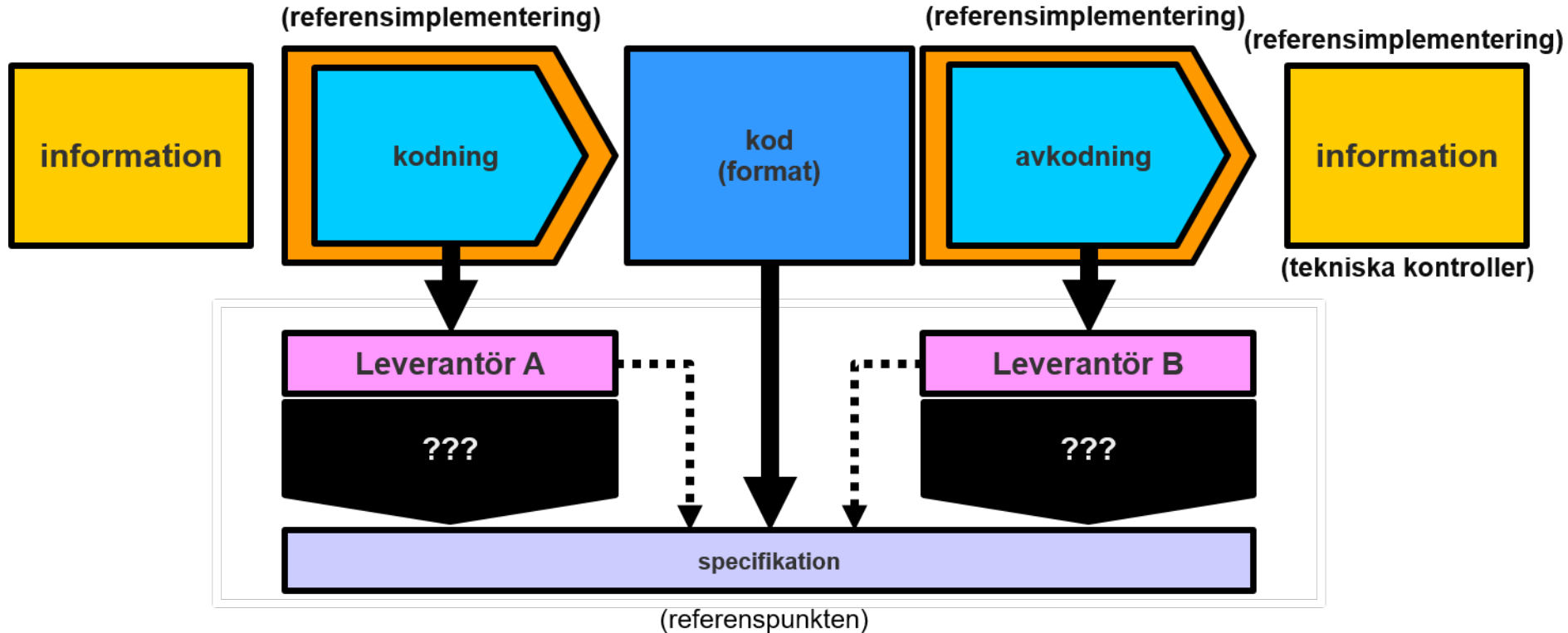
2. Vad är FormatE? – Avgränsning



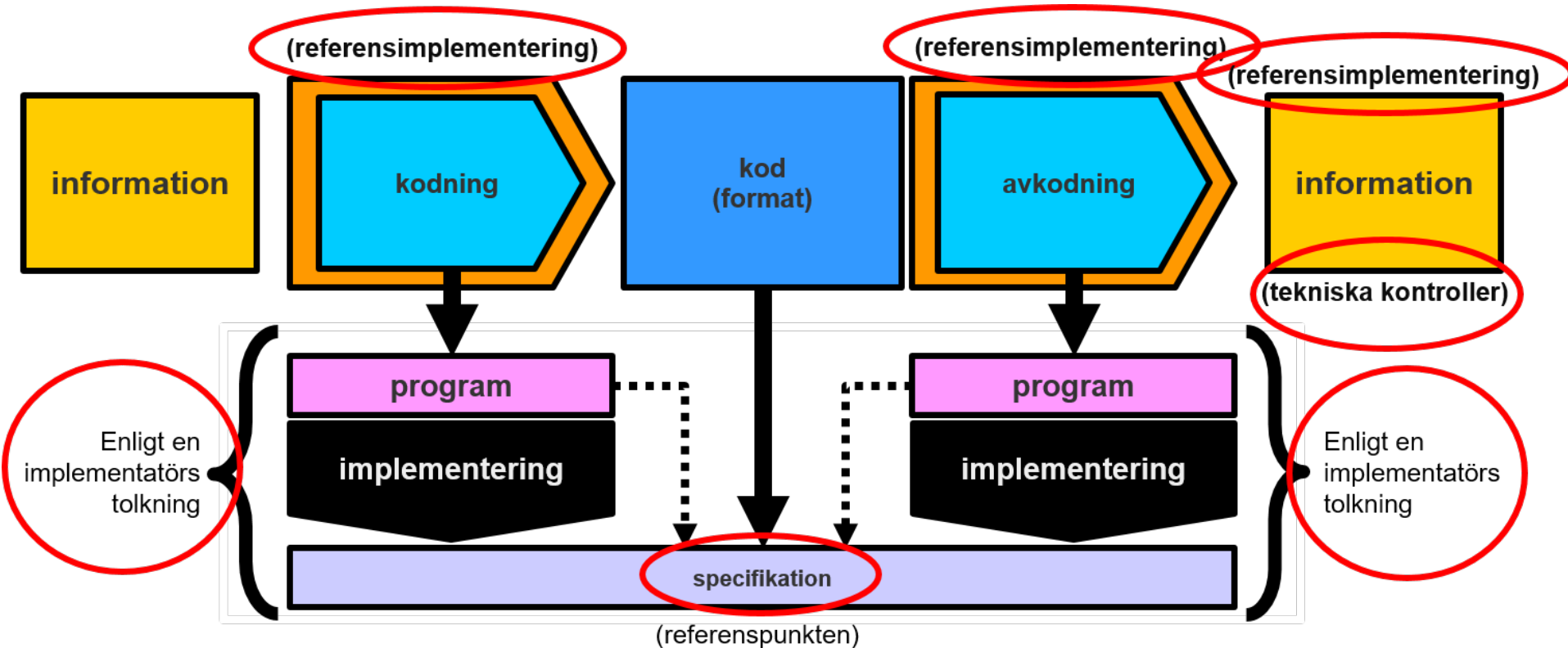
2. Vad är FormatE? – Modellen



2. Vad är FormatE? – Metoder för mätbarhet



2. Vad är FormatE? – Strategiska mål



2. Så vad är FormatE...?

Till exempel

- En metodlära (källor och deras tolkning – standardisera implementeringar)
- En ”strategi” (strategiska mål)
- En ”taktik” (krav vid framställning)
 - En omvärldsbevakning, dokumentation (specifikationer)
 - En normering (vägledningar och rättsregler)
- En metod för mätning (teknisk kontroll)
- Modeller för bedömning och värdering (gallring, öppenhet, val av materiel och metoder)
- Ett kunskapsarkiv



2. Vad är FormatE? – Normeringen

- Tekniska krav – interoperabilitet – vad krävs för koda, avkoda (framställa).
 - Till exempel, öppna med olika program, konvertera till, från olika format
- Arkivrättsliga krav – vad krävs för bevarande (använda och hantera i arkivet)
 - Till exempel, läsa och förstå innehållet
- Funktionellt skick – vad som är av betydelse för 3 § arkivlagen (värdera)
 - 3 § arkivlagen och verksamhetens strategi för bevarande
 - gallra helt
 - gallra i delar (t.ex. webbsidor till PDF/A, 3D-modeller till bilder)
 - bevara i ursprungligt skick (t.ex. elektroniska fakturor, tekniska ritningar)

2. Vad är FormatE? – De tekniska kraven

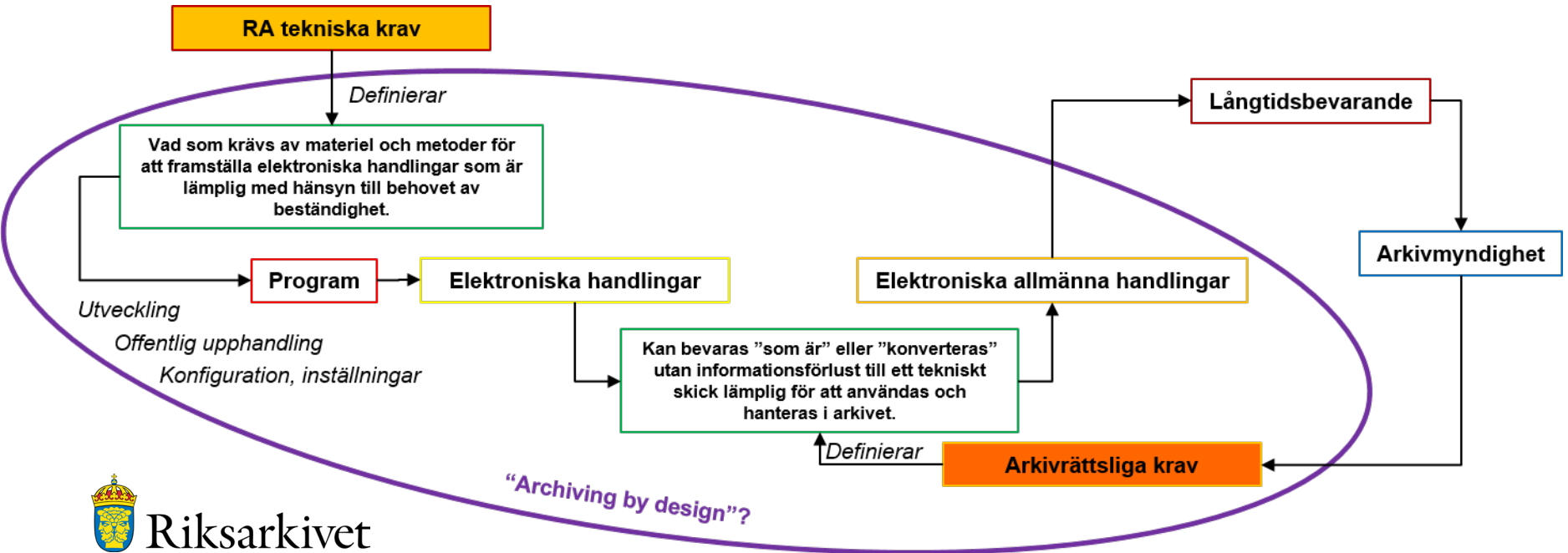
- Vad som är lämpligt med hänsyn till behovet av beständighet
- Kraven tvingar inte verksamheter att använda en eller annan specifikation men vilka alternativ är lämpliga för olika fall. Till exempel,
 - Ett allmänt krav är att kryptering är *inte lämpligt för beständigheten* av en elektronisk handling
 - Varje verksamhet måste fortfarande överväga om *användningen* av kryptering överväger behovet av beständighet
 - Arkivmyndigheten får sedan kräva dekryptering vid överlämnande.

2. Vad är FormatE? – De arkivrättsliga kraven

- Krav på framställning utesluter inte krav på överlämnande
 - Framställ enligt arkivrättsliga krav – Kostnaderna hamnar på arkivmyndighet
 - Konvertera vid överlämning till arkivmyndighet – Kostnaderna hamnar hos överlämnaren
- Arkivmyndigheten tillsammans med verksamheterna får närmare klargöra vad som ska gälla

2. Vad är FormatE? – Målbilden

- Ett skifte av fokus – från krav vid överlämnande till krav vid framställning



-
- 3. Vad är problemet?
-
-
-
-
-
-

Exempel på vad vi försöker uppnå med FormatE

3. Vad är problemet? – ”Det digitaliserade samhället”

- All ”information” i ett samhället kommer att representeras av ett eller annat format.
 - Bioteknik, DNA-teknik är på väg...
- Jämför
 - Uttalanden som att ”data, information är den nya oljan”
 - Riksarkivets uttalande mål om att ha ett helhetsperspektiv på informationsförsörjningen i samhället

3. Vad är problemet? – ”Det digitaliserade samhället”

FormatE – *Kartlägga specifikationer* – primärkällan!

1. Få ordning på källunderlaget
 - Grundspekifikationer, utökningar av dem, och tillägg till dem?
 - Specifikationens tillämpningsområden?
2. Dokumentera tolkningar – sekundära källor!
 - Jämför bevara [eller förlorad] kunskap inom andra områden

3. Vad är problemet? – ”Det digitaliserade samhället”

FormatE – *Metodläran*

- Ett sätt att hantera
 - ”tekniken går snabbt”
 - format blir ”gamla”, ”föråldrade”
- Bidra till konsekvent implementering
- Jämför *juridisk metodlära*

3. Vad är problemet? – PDF 1.5/A-1 (ISO 19005-1)

ANSI X3.4, Information Systems — Coded Character Sets — 7-Bit American National Standard Code for Information Interchange (7-Bit ASCII) ■ ECMA-6, 7-Bit coded Character Set. ■ ISO 639-1, Codes for the representation of names of languages — Part 1: Alpha-2 code ■ ISO 2108, Information and documentation — International standard book number (ISBN) ■ ISO 3166-1, Codes for the representation of names of countries and their subdivisions — Part 1: Country codes ■ ISO/IEC 10646-1:2000/Amd 1:2002, Amendment 1: Mathematical symbols and other characters ■ ISO/IEC 10646-2:2001, Information technology — Universal Multiple-Octet Coded Character Set (UCS) — Part 2: Supplementary Planes ■ ISO/IEC 14492:2001, Information technology — Lossy/lossless coding of bi-level images ■ Equivalent to ITU-T Recommendation T.88, Information technology — Coded representation of picture and audio information — Lossy/lossless coding of bi-level images ■ ISO 15489-1:2001, Information and documentation — Records management — Part 1: General ■ ISO/TR 15801, Electronic imaging — Information stored electronically — Recommendations for trustworthiness and reliability ■ ISO 15930-6, Graphic technology — Prepress digital data exchange using PDF — Part 6: Complete exchange of printing data suitable for colour-managed workflows using PDF 1.4 (PDF/X-3) ■ ISO/TR 18492, Long-term preservation of electronic document-based information ■ ISO 18509-1, Electronic archival storage — Specifications relative to the design and operation of information processing systems in view of ensuring the storage and integrity on recordings stored in these systems — Part 1: Long term access strategy ■ ISO 18509-2, Electronic archival storage — Specifications relative to the design and operation of information processing systems in view of ensuring the storage and integrity on recordings stored in these systems — Part 2: Technical specifications ■ ITU-T Recommendation T.4, Standardization of Group 3 facsimile terminals for document transmission ■ ITU-T Recommendation T.563, Terminal characteristics for Group 4 facsimile apparatus ■ Language Tags, IANA. ■ Namespaces in XML 1.1, W3C Recommendation, February 4, 2004. ■ PDF Reference: Adobe Portable Document Format, Version 1.5, Adobe Systems Incorporated — 4th edition. ■ RFC 1950, ZLIB Compressed Data Format Specification version 3.3, May 1996. ■ RFC 1951, DEFLATE Compressed Data Format Specification version 1.3, May 1996. ■ The Unicode Standard, Unicode Consortium. ■ Unicode Standard Annex #15, Unicode Normalization Forms, Unicode Consortium, 17 April 2003. ■ ISO 639-2, Codes for the representation of names of languages — Part 2: Alpha-3 code

PDF Reference 1.5

Technical Notes: 22st ■ Litteratur: 13st ■ Adobe Type 1 Font Format. ■ OPI: Open Prepress Interface Specification 1.3. ■ PDF Public-Key Digital Signature and Encryption Specification. ■ PostScript Language Reference, Third Edition, Addison-Wesley, Reading, MA, 1999. ■ XMP: Extensible Metadata Platform. ■ Apple Computer, Inc., TrueType Reference Manual. ■ ICC.1:1998-09, File Format for Color Profiles, and Document ICC.1A:1999-04, Addendum 2 to Specification ICC.1:1998-09 ■ ICC Characterization Data Registry ■ CIP4 JDF Specification, Version 1.0. ■ International Electrotechnical Commission (IEC), IEC/3WD 61966-2.1, Colour Measurement and Management in Multimedia Systems and Equipment, Part 2.1: Default RGB Colour Space—sRGB. ■ ISO 639, Codes for the Representation of Names of Languages ■ ISO 3166, Codes for the Representation of Names of Countries and Their Sub-divisions ■ ISO/IEC 8824-1, Abstract Syntax Notation One (ASN.1): Specification of Basic Notation ■ ISO/IEC 10918-1, Digital Compression and Coding of Continuous-Tone Still Images (informally known as the JPEG standard, for the Joint Photographic Experts Group, the ISO group that developed the standard) ■ International Telecommunication Union (ITU), Recommendations T.4 and T.6. ■ RFC 1321, The MD5 Message-Digest Algorithm ■ RFC 1738, Uniform Resource Locators ■ RFC 1766, Tags for the Identification of Languages ■ RFC 1808, Relative Uniform Resource Locators ■ RFC 1866, Hypertext Markup Language 2.0 Proposed Standard ■ RFC 2045, Multipurpose Internet Mail Extensions (MIME) Part One: Format of Internet Message Bodies ■ RFC 2046, Multipurpose Internet Mail Extensions (MIME) Part Two: Media Types ■ RFC 2068, Hypertext Transfer Protocol—HTTP/1.1 ■ RFC 2083, PNG (Portable Network Graphics) Specification, Version 1.0 ■ Microsoft Corporation, TrueType 1.0 Font Files Technical Specification. ■ Netscape Communications Corporation, Client-Side JavaScript Reference. ■ Unicode Standard Annex #9, The Bidirectional Algorithm, Version 3.1.0, and Unicode Standard Annex #14, Line Breaking Properties, Version 3.1.0. ■ Extensible Markup Language (XML) 1.0 ■ Extensible Stylesheet Language (XSL) 1.0 ■ Web Content Accessibility Guidelines 1.0

3. Vad är problemet? – "MPEG-4" (ISO/IEC 14496)

- Register över specifikationer – *identifierare för specifikationer*
- [ISO/IEC 14496-3] Adif (Audio Data Interchange Format)
- [ISO/IEC 14496-3] AAC
- [ISO/IEC 14496-16] AFX (Animation Framework eXtension)
- [ISO/IEC 14496-25] 3D Graphics Compression Model
- [ISO/IEC 14496-23] SMR (Symbolic Music Representation)
- [ISO/IEC 14496-3] ALS (Audio Lossless Coding)
- [ISO/IEC 14496-11, -21] "MPEG-J"
- [ISO/IEC 14496-17] STF (Streaming text format)
- [ISO/IEC 14496-20] LAsER
- [ISO/IEC 14496-20] SAF (Simple Aggregation Format)
- [ISO/IEC 14496-1] MP4 file format (version 1)
- [ISO/IEC 14496-14] MP4 file format (version 2)
- [ISO/IEC 14496-15] NAL (Network Abstraction Layer)
- [ISO/IEC 14496-2] Visual
- [ISO/IEC 14496-10, ITU-T H.264] AVC
- [ISO/IEC 14496-29] WVC (Web Video Coding)
- [ISO/IEC 14496-33] IVC (Internet Video Coding)
- [ISO/IEC 14496-28] CFR (Composite Font Representation)
- [ISO/IEC 14496-22] OFF (Open Font Format)

3. Vad är problemet? – Formatet är allt

- En uppfattning om att formatet ”i sig” är tillräckligt för bevarande
- Till exempel
 - ”arkivformat”, ”beständiga format” eller ”arkivbeständiga format”
 - ”PDF/A”, ”HTML”, ”XML”, ”PNG”, ”WAVE”

3. Vad är problemet? – Formatet är allt

FormatE – format *och program* – tekniskt skick

1. Identifiera lämpliga referensimplementeringar
 - Sekundär källor men auktoritära tolkningar!
 - Till exempel, programbibliotek, kodbibliotek
2. Rekommendera
 - Vilka uppfyller tekniska krav?

3. Vad är problemet? – Inte ett problem förrän ett problem

- Vanligtvis när handlingen ska användas och hanteras

Vi har just nu på att utvärdera Acrobat Pro för konvertering av mail och kontorsdokument till PDF/A.
Vi får ofta felmeddelanden i Preflight med att det inte gick att göra en godkänd PDF/A av filen.
Så därför undrar jag om ni har eller kommer att ha någon kurs i hur man framställer eller konverterar en
hur man använder Preflight och hanterar felmeddelanden i Preflight.
Eller om ni känner till något företag som har kurser i detta

- Konsekvenser vid masshantering?

ghet att föreläsa om din utre
Digitalt eller på plats på Bru

3. Vad är problemet? – Inte ett problem förrän ett problem

FormatE – *tekniska kontroller*

1. Kartlägg och analysera befintliga metoder
 - Vilka tekniska krav kan de kontrollera? (format)
 - Vad kontrollerar dem? (materiell, formell, policy, form-funktion)
2. Eventuellt ta fram nya metoder
 - Till exempel, Preforma

3. Vad är problemet? – Val av lämpliga format

- Vad gör ett ”arkivformat” mer eller mindre lämpligt?
- Det finns ett antal vanligt förekommande kriterier. Till exempel,
 - särskilt dem från kongressbiblioteket i USA, eller
 - diverse utredningar och arbeten, vanligtvis från nationella arkiv

3. Vad är problemet? – Val av lämpliga format

- Bedömningen av kriterierna är emellertid godtyckligt. Till exempel:
 - ”Bred användning” – vad betyder det?
 - Jämför ”MPEG-4”, Flash, Tiff
 - Gäller samma villkor för alla tillämpningsområden? – jämför format för media, nätverk, kommunikation, kontor, industri
 - ”Kunskap och kompetens” eller ”tillgång till program – jämför Matroska+FFV1, DocX och ODT, PDF/A-1 -2 -3 nivå ”b”, ”u” och ”a”
- Bedömningen kan väsentligt skilja sig åt mellan ”arkivverksamheter”
 - Jämför International Comparison of Recommended File Formats
 - <https://openpreservation.org/news/new-community-resource-international-comparison-of-recommended-file-formats/>



3. Vad är problemet? – Val av lämpliga format

FormatE – kriterier för val av materiel och metoder

- Att mäta kriterierna är svårt; den utredningen fortsätter
- Fokus på vad som med tiden krävs av verksamheten för att koda, avkoda
 - *Konsekvenserna av val av materiel och metoder*
 - *Går det att koda, avkoda konsekvent?*

3. Vad är problemet? – Att koda, avkoda konsekvent

- Vad är syftet med kriterierna för ”val av format”?
 - Att lagra formaten? (”binärt bevarande”)
 - Att även kunna återge formaten?
 - Hur återges formaten?
 - Kan formaten återges?
 - Vad kan återges av formaten?
 - Vad ska återges av formaten?

3. Vad är problemet? – Att koda, avkoda konsekvent

- Konsekvent framställning och återgivning
- Tolkning och implementering
 - Ett olösligt problem?
- Principiellt samma problem i andra sammanhang
 - Digitala signaturer – olika policy vid validering
 - Ossifikation – nätverksprotokoll
 - E-fakturor – leverantörsspecifika anpassningar

3. Vad är problemet? – Att koda, avkoda konsekvent

- Val av specifikationer
- Val vid implementering [av specifikation]
- Val av program eller källkod
- Val vid kompilering av källkod till program
- Val av inställningar

3. Vad är problemet? – Att koda, avkoda konsekvent

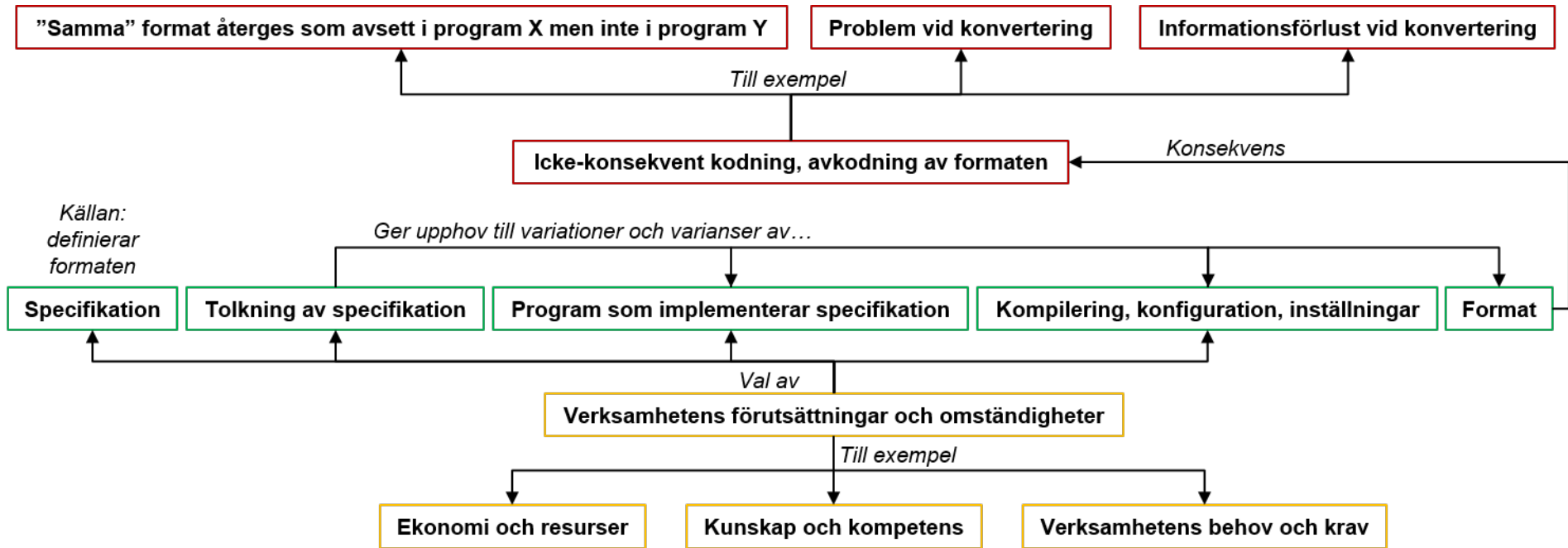
Verksamhetens förutsättningar och omständigheter:

- Ekonomi och resurser
- Kunskap och kompetens
- Rättsliga och interna regler

Till exempel:

- Vissa verksamheter föredrar att arbeta i XML, andra i PDF
- Arkivmyndigheter kan hantera vissa format men inte andra

3. Vad är problemet? – Att koda, avkoda konsekvent



Frågor?