|  |  |  |
| --- | --- | --- |
|  | | Memorandum |
| |  |  | | --- | --- | | Title: | Guide to P1 Migration Assessment | | To: | New Preservation Concept | | From: | Asbjørn Skødt | |  |  | |  |  | |  |  | |  | Date: 18. november 2020  Case no.: 19/06296 |

# 1. Introduction

This step is the most important step in the Concept Model for Development of Preservation Plans. The step is important because it is the results of this assessment, which determines if the investigated content type can be migrated to one or more of the existing preservation formats with an acceptable loss of quality measured in loss of significant properties.

The step P1 Migration Assessment leads to one of the following recommendations:

1. **Continue to P2 Format Assessment** if the content type cannot migrate with an acceptable loss of data quality to one or more of your archive’s existing preservation formats.
2. **Continue to P5 Preservation Plan** if the content type can migrate with an acceptable loss of data quality to one or more of your archive’s preservation formats.

Migration is in this context defined as the conversion of data between file formats.

# 2. Method

The migration assessment applies the InSPECT (Investigating Significant Properties of Electronic Content) framework for identification of significant properties. Gareth Knight developed the framework in 2009 through JISC funds and it was hosted as a cooperation between The Centre for e-Research at King’s College, London, and the British National Archives.

In summary, the InSPECT framework is a method to assessing the significant properties of file formats’ structure and usage. The framework enables archives to make documented preservation plans and actions supporting the technical mapping and validation of archival content and thus reinforcing authenticity of archived data. In the context of InSPECT, authenticity is considered a relative term, which defines the level of preservation through an archive’s ability to preserve significant properties when migrating data.

Knight uses the following definition of significant properties:

*”The characteristics of digital objects that must be preserved over time in order to ensure the continued accessibility, usability, and meaning of the objects and their capacity to be accepted as evidence of what they purport to record.”*

The framework utilizes to other methods from the field of digital preservation and engineering. The first one is the recognized and highly adopted OAIS reference model, which defines a group of stakeholders as a “designated community”.

Knights metode læner sig op af to eksisterende metoder inden for hhv. digital bevaring og ingeniørvidenskab. Den første er den anerkendte og meget udbredte OAIS-model, der bl.a. definerer et brugerfællesskab (”designated community”), dvs. interessenter etableres som entiteter for arkivet, og det betyder, der skal etableres en passende planer for håndtering af interessenterne. Den anden metode er den såkaldte *Function-Behaviour-Structure (FBS)*, som er en procesunderstøttende designmetode for ingeniører, der leverer en metode til at samarbejde med brugere om en forståelse af deres behov.

Der henvises til notatet ”InSPECT framework til vurdering af signifikante egenskaber v1.0” sagsdok. 19/06296-11 for yderligere information om metoden.

## 2.1. Metodiske fravigelser

Migreringsvurderingen anvender ikke InSPECT slavisk, idet vi gennem erfaring har fundet ud af, at der er aspekter, som der ikke entydigt er formålstjenstlige. Det gælder de følgende:

* Det er ikke et krav, at alle egenskaber skal identificeres og redegøres for, fordi det kan være for omfattende en opgave ved komplekse indholdstyper. I stedet skal vi fra starten koncentrere vurderingen om de egenskaber, som vi har en formodning om er vigtige for indholdstypen. Det er de egenskaber, hvis vigtighed vi skal opnå interessenternes validering af. Det er afgørende, at vi får kortlagt alle vigtige egenskaber dog.
* Interessentanalysen har et skridt, der omhandler tilknytning af interessentens acceptable grænseværdier, f.eks. konkrete min-max værdier for billedopløsning eller bitdybde. Den viden kan bruges til specificering og validering af datakvalitet. Det kan dog være svært for interviewpersoner entydigt at definere værdier. Det er derfor ikke et krav men en anbefaling, at interviewpersoner skal definere deres grænseværdier.

# 3. Udførsel

I de følgende afsnit beskrives, hvordan migreringsvurderingen i praksis udføres, og hvilke produkter der forventes. Se nedenstående oversigtsillustration af proces og produkter.

Produkter

Interessentanalyse

Teknisk analyse

## 3.1. Teknisk analyse af vigtige egenskaber

De identificerede filformater i det foregående procestrin P0 Foranalyse er objekterne for den tekniske analyse. Analysen af filformaternes tekniske struktur kan ligeledes tage udgangspunkt i den foreløbige tabel over vigtige egenskaber, der er identificeret i det foregående procestrin P0 Foranalyse.

Analysen kan involvere indsamling af specifikationer, test med karakteriseringsværktøjer, konkret brug af data, research på internettet og samtale med kollegaer. Det kan i den sammenhæng være nødvendigt at fremskaffe eller producere testdata evt. i dialog med interessenter.

De identificerede vigtige egenskaber anføres i en tabel. Tabellen over vigtige egenskaber skal leveres i regnearket ”Tabel\_Vigtige egenskaber”.

Tabellen har følgende kolonner, som skal udfyldes:

* **Kategori:** Der vælges en kategori, som den vigtige egenskab bedst passer i. Valgmuligheder er ”Indhold”, ”Kontekst”, ”Visning”, ”Struktur” og ”Opførsel”[[1]](#footnote-1).
* **Underkategori:** Fri angivelse af egenskabens underkategori f.eks. ”Filtrering og sortering”, hvormed alle egenskaber forbundet til filtrering og sortering kan fremsøges og valideres.
* **ID:** Der angives et fortløbende id f.eks. 1, 2, 3 osv.
* **Navn:** Den vigtige egenskab skal angives med et navn, der er sigende for egenskabens indhold.
* **Beskrivelse:** Den vigtige egenskab skal kort beskrives.
* **Konverteringstab:** Tabes den vigtige egenskab ved konvertering til et eller flere af Rigsarkivets eksisterende bevaringsformater? Enten ja eller nej.
* **Interessentvalideret:** Har en eller flere interessenter tilkendegivet egenskaben som vigtig? Enten ja eller nej.
* **Interessentbemærkninger:** De evt. bemærkninger, som interessenten har til egenskaben herunder f.eks. angivelse af grænseværdier.

Regnearket har ligeledes to andre ark: ”Dataanalyse” og ”Dataformater”.

* **Dataanalyse** anvendes til at kvantificere og overskueliggøre, hvor mange vigtige egenskaber tabes ved konvertering, og hvor mange af disse er interessentvalideret. Der anvendes en pivottabel til fremvisningen.
* **Dataformater** er en tabel, hvor de vigtige egenskaber skal mappes op imod de filformater, som er udvalgt i P0 Foranalyse til vurdering i P2 Formatvurdering. Her registreres ”Ja/Nej” for hver vigtig egenskab ved hvert filformat. Formålet er at skabe en registrering af, hvilke filformater bevarer flest mulige vigtige egenskaber. Den viden skal anvendes i formatvurderingen.

## 3.2. Interessentanalyse

Interessentanalysen har til formål at tilvejebringe viden om interessenters produktion og/eller brug af indholdstypens data. InSPECT-metoden ser, at *”significance is in the eyes of the beholder”*, hvorfor analysen skal bruges til at kvalificere betydningen af vigtige egenskaber, således at det ikke bliver en akademisk skrivebordsøvelse for arkivets egne medarbejdere at forestille sig, hvad der efterspørges af en bruger om f.eks. 50 år. En god indledning til interviewet er netop at fortælle interviewpersonen, at interviewet handler om at tilvejebringe en forståelse for, hvordan de ville bruge data, hvis de forestillede sig at være en fremtidig bruger af dataene.

### 3.2.1. Identificering af interessenter

Interessentanalysen involverer typisk interviews med 3-5 interessenter, men der er ingen nedre eller øvre grænser. Det anbefales dog at interview mindst en kommune og en statslig myndighed. Interessenter er typisk dataproducenter (oftest myndigheder), andre arkiver, brugere og konsulenter.

Identificering af mulige interviewpersoner tager udgangspunkt i den identificering, der allerede er foretaget i det foregående procestrin, P0 Foranalyse, men der kan identificeres yderligere, og der skal træffes beslutning, om hvilke der skal kontaktes med henblik på et interview. Et interview afholdes med de personer, som interessenten finder relevante for deltagelse, men en erfaring er, at jo færre, der deltager i interviewet, jo bedre.

Migreringsvurderingens notat skal indeholde en tabel over, hvilke interessenter (interviewpersoner) er afholdt interviews med, hvilke er kontaktet uden svar eller interview er afvist samt kontaktinformation.

### 3.2.2. Udvikling af spørgeramme

Der skal udarbejdes en eller flere spørgerammer til brug for interviews. Det er en god idé at fremsende spørgerammen til interviewperson(er) på forhånd, så de har mulighed for at forberede sig til interviewet og dermed være i stand til at besvare spørgsmålene bedre og mere præcist. Spørgerammen tilpasses til interviewsituationen, således at kun målrettede spørgsmål til interviewpersonen stilles.

Spørgsmålene skal formuleres pædagogisk, men de skal formuleres på en måde, således at arkivet opnår viden, om hvilke filformater bruges (og til hvilke formål: Skabelse, dataudveksling og/eller (gen)anvendelse af data), hvilke egenskaber interviewpersonen finder vigtige, og hvilke grænseværdier for de vigtige egenskaber er acceptable, hvis vedkommende skulle genanvende dataene i fremtiden.

Det er også i forlængelse af spørgerammen, at der skal stilles spørgsmål, om interviewpersonen kan levere repræsentative eksempler på de data, som interviewet omhandler. Det er særligt relevant for producenter af data såsom myndigheder.

### 3.2.3. Afholdelse af interviews

Interviews kan afholdes med personligt fremmøde hos interviewpersonen eller gennem internetmøde. Interviews anbefales at vare 1-2 timer.

Der skal skrives referat fra alle interviews, og det kan lette arbejdet med efterfølgende at sammenstille brug med vigtige egenskaber, hvis der løbende under interviewet afkrydses i tabellen for vigtige egenskaber, når de verbalt bliver brugervalideret. Derfor er det altid en god idé at deltage med flere medarbejdere i interviewet, således at arbejdsbyrden fordeles, og samtalen lettes.

## 3.3. Sammenstilling af vigtige egenskaber og interessenters brug

Sammenstilling af vigtige egenskaber og interessenters brug har til formål at be- eller afkræfte vores formodning, om at en egenskab er vigtig. Det sker ved at forbinde viden fra de to foregående analyser af vigtige egenskaber og interessenters brug. Vi opnår herved en ”virkelighedsvalidering” af vigtige egenskaber, frem for at arkivet skal gisne, om hvad der er vigtigt at bevare, når arkivet konverterer data.

Skridtet udføres ved at analysere svarene fra interviews og kortlægge dem op mod tabellen med vigtige egenskaber og evt. angive nye vigtige egenskaber, som interviews har bidraget med viden om. Der angives ”Ja/Nej” i kolonnen ”Interessentvalideret” i regnearket ”Tabel\_Vigtige egenskaber”.

# 4. Produkter

Migreringsvurderingen har tre leverancer:

* Et notat, der beskriver vurderingens forløb og udfald. Notatet skal rumme en tabel over identificerede interessenter jf. 3.2.1. Notatet skal ligeledes rumme en tabel over identificerede filformater, der anvendes som grundlag for migreringsvurderingen og til vurdering i procestrinet P2 Formatvurdering, eller en anvisning af, hvilke eksisterende bevaringsformater kan bevare indholdet til specificering i procestrinet P5 Bevaringsplan. Tabellen med filformater skal indeholde information om filformatet anvendes til skabelse, dataudveksling og/eller (gen)anvendelse af data.
* En tabel i et regneark, der oplister identificerede vigtige egenskaber samt hvorvidt egenskaben tabes ved konvertering til et eller flere af Rigsarkivets eksisterende bevaringsformater, og hvorvidt egenskaben er be- eller afkræftet af en interessent. Regnearket opmærker ligeledes, hvilke filformater besidder hvilke vigtige egenskaber.
* Eksempeldata leveret af interessenterne. Data skal helst udgøre et repræsentativt udsnit af interessentens data for den undersøgte indholdstype og/eller filformat.

Migreringsvurderingen skal samlet levere en anbefaling til, hvorvidt et eller flere af Rigsarkivets bevaringsformater på tilfredsstillende vis kan bevare indholdstypens eller det konkrete formats vigtige egenskaber. Hvis ikke skal formatvurderingen af et eller flere nye bevaringsformater igangsættes. Migreringsvurderingen leveres som en anbefaling til ledelsesmæssig stillingtagen, om hvad der skal ske.

1. **Content**: Information content within the Information Object. For example, text, still and moving images, audio, and other intellectual productions. Examples: duration, character count.

   **Context**: Any information that describes the environment in which the Content was created or that affect its intended meaning. Examples: Creator name, date of creation.

   **Rendering**: Any information that contributes to the re-creation of the performance. For example, font type, colour and size, bit depth.

   **Structure**: Information that describes the extrinsic or intrinsic relationship between two or more types of content, as required to reconstruct the performance. E.g. e-mail attachments.

   **Behaviour**: Properties that indicate the method in which content interacts with other stimuli. For example, hyperlinks. [↑](#footnote-ref-1)