



Noark

The Norwegian Records Management Standard

- The format of repository transfers from Noark systems
- Experiences with Noark transfers since the early 1990's

Prague 22 April 2009

Hans Fredrik Berg

The National Archives of Norway

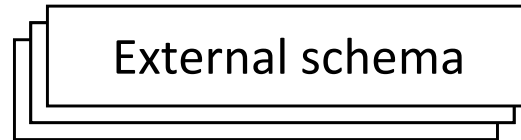
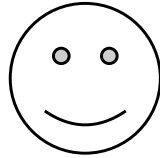
The Noark standard

- A standard for ERM systems in Norwegian government and municipal agencies since 1984
- A mandatory standard since 1999 (also for registries embedded in specialized business applications)
 - Approximately all stand-alone ERM systems in public agencies are Noark systems certified by The National Archives
- Version 1-3 (1984-1994): A registry systems standard
- Version 4 (1999) and after: An ERM standard including electronic archiving of documents
- Version 5 (2008): MoReq2 compatibility process started

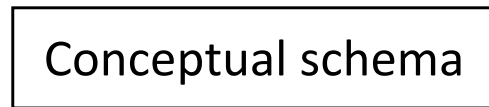
Noark version 1-3 (1984-1994)

- A standard for registry systems ("journal systems")
 - Continuous registration of reference information (metadata) on documents received and sent by agencies
 - Including references to the filing of paper documents to enable management of physical archives
 - From version 3 (1994) Noark systems might also include references to electronic work copies of paper documents
- Two levels of standardization
 - Functional requirements
 - Transfer format (but samples were not tested and certified by The National Archives)

Three-schema architecture



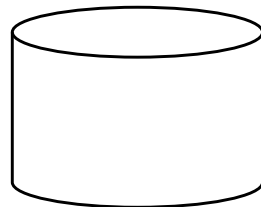
user view



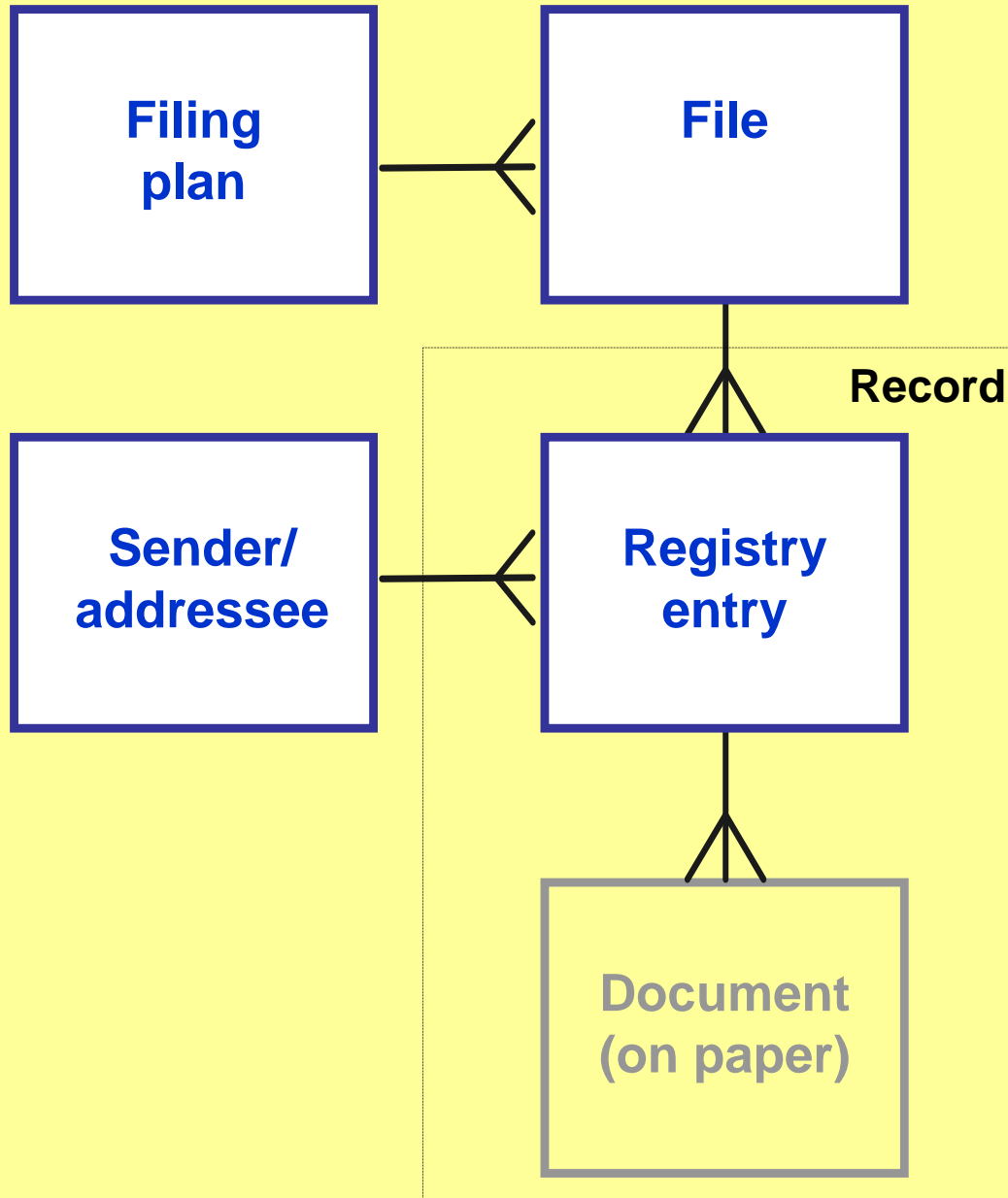
“business” logics



data definition



Noark version 1-3 (1984-1994) data model



Transfers to The National Archives had to conform to this model

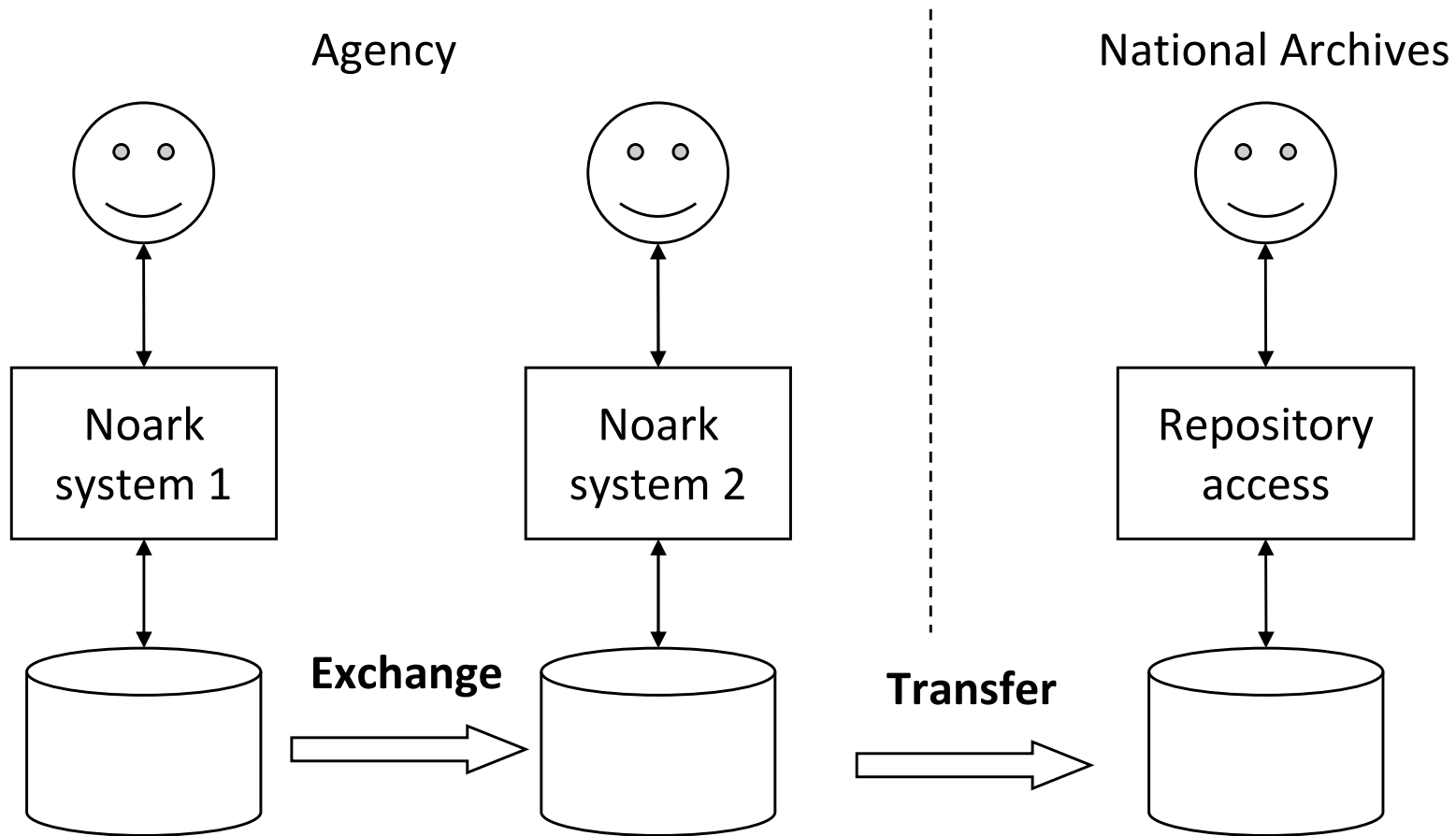
Each table had to contain specific fields with specific names, data types, lengths, etc.

The native format of Noark systems might be different (and more complex), but data had to be converted to this structure when transferred

The transfer format of Noark version 1-3

- Exported data extracts from 3 or 4 tables
 - 40 columns (metadata) included
 - Table content exported as flat files in fixed (length) format
- Purpose: To enable regeneration of technical and logical structure
 - The Noark 1-3 model did not distinguish between these two structures
 - The four tables defined were not required to be implemented as the native (file or database) structure of Noark-systems
- One common format was used for repository transfer and data exchange between Noark systems
- Repository access: Dependent on Noark 1-3 viewers (or "survived" Noark 1-3 applications)

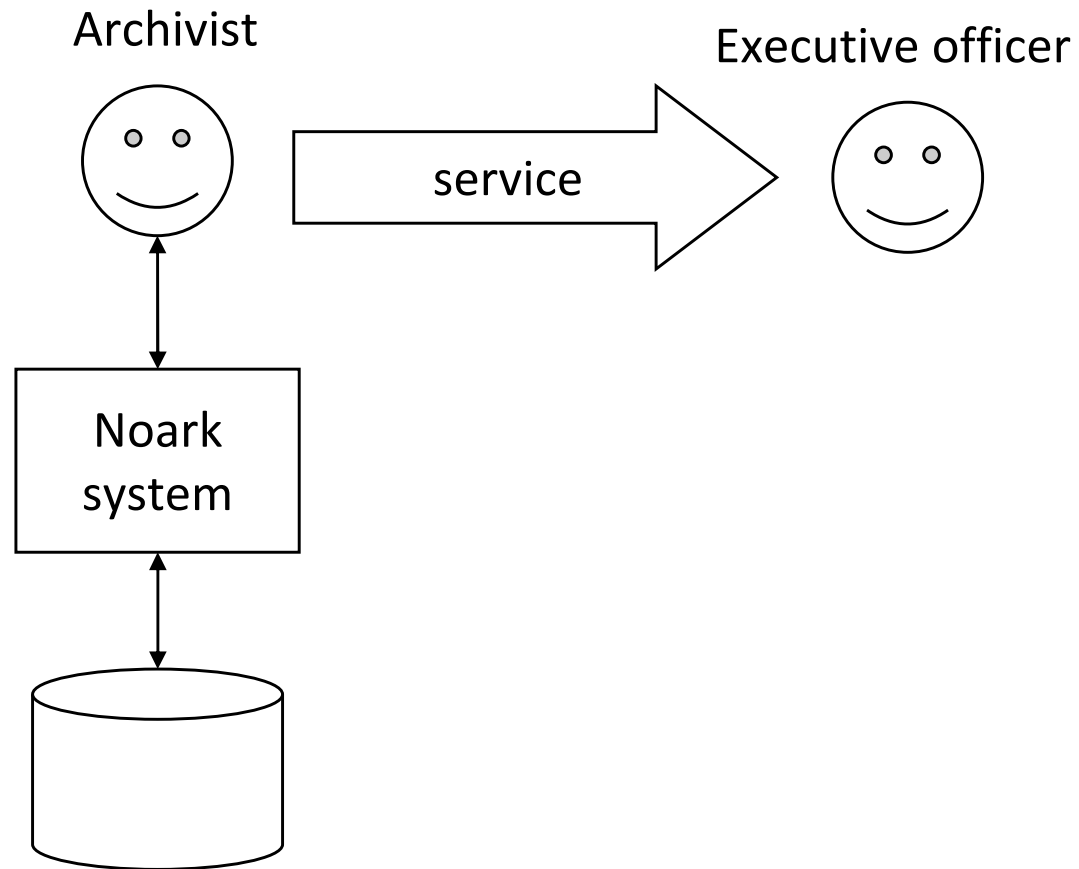
Exchange and transfer



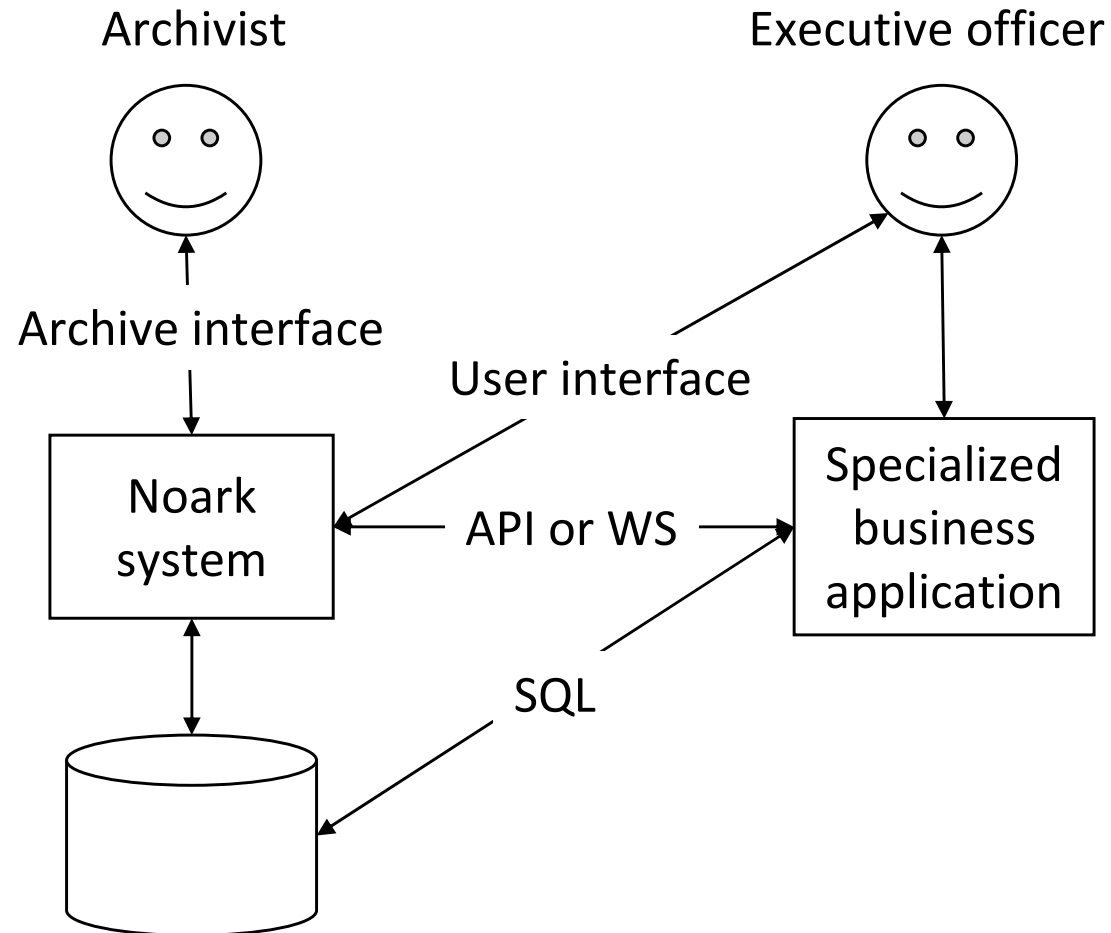
Noark version 4 (1999)

- Specification of an ERM system
 - Technical structure specified in great detail (Noark-4, part 2)
 - 95 tables with 885 columns
- Two levels of standardization
 - Functional requirements
 - Transfer format (samples are tested by The National Archives during the certification process)
 - *But: it is not mandatory to implement the technical structure of Noark-4 as the application's native (internal) structure*
 - *What is mandatory, is the capacity to output transfers according to the structure defined in Noark-4, part 2.*

Typical use of Noark 1-3



Typical use of Noark 4



The transfer format of Noark-4

- Data are exported from 40-60 tables
 - Data are exported in XML format
 - A DTD is defined for each table
- The transfer format is basically the same as the format for exchange between Noark-4 applications
- Like Noark v. 1-3, the purpose is to enable regeneration of technical structure
- Repository access
 - Database applications (Modified Noark-4 systems)
 - Database report generators
 - Database querying (SQL)

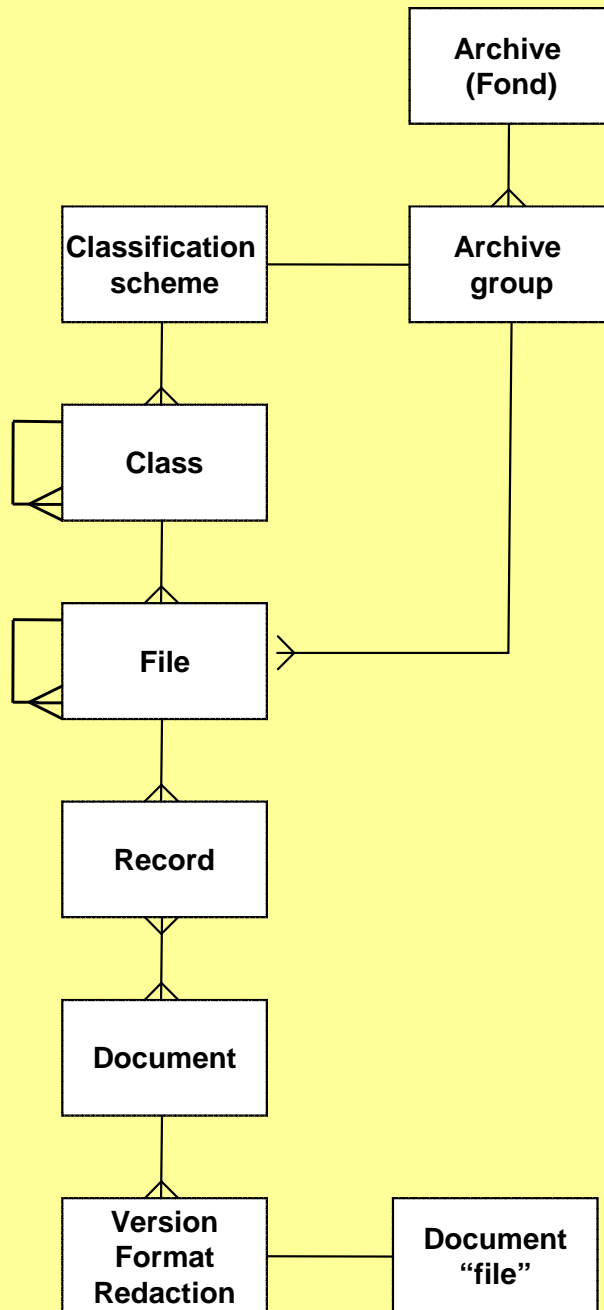
Problems with Noark-4 transfers

- Data in table extracts are frequently inconsistent
 - Referential integrity rules are violated
 - Duplicated primary key values occur in tables
 - The transfer model is dependent on referential integrity
- This type of problems was first observed in Noark-3 transfers. What may cause the problems?
 - Conversions from a native database structure
 - Earlier conversions of data from a "liberal" Noark-3 system to a Noark-4 system
 - Problems in fulfilling the requirements when selecting closed files for transfer (the principle of "periodization" in Noark)
 - Complexity in general
- Conclusion: A need for simplified transfer structures

Noark version 5 (2008)

- Specification of an ERM standard
 - Logical structure specified
- Two levels of standardization
 - Functional requirements
 - *Very few mandatory core requirements*
 - *Optional requirements for integration with other tools to build complete stand-alone ERMS' (like Noark-4)*
 - *Optional requirements for embedding in business applications*
 - Transfer format (samples will be tested by The National Archives during the certification process)
- Focus on the records rather than the applications

Noark 5 data model



Archive: Possible to include several archives (created by different agencies) in one system

Archive group: Division of the archive into separate periods of records, typically 5 years each

File: Arbitrary number of levels of sub-files

Component: used to handle different versions, redactions and formats of the same document

Transfer: Each document should be transferred as one file (bit stream). The National Archives accept the following document formats:

- Text (UTF-8 or ISO 8859)
- TIFF v. 6
- XML (ODF or OOXML?)
- PDF/A

The transfer format of Noark 5 (2008)

- Transfers are radically simplified - and more robust
 - Confined to 8 archival units (metadata objects)
 - 175 metadata in one hierarchical XML file as defined in an XML Schema
- Archival information is liberated from the technical structure
 - By focusing on archival records, preservation can be confined to the logical structure of the records themselves
 - Identification of the 8 archival units is not based on the original database primary key values
 - All codes employed by the application system are substituted by the texts represented by the codes

Consequences

- Completely different language (XML vs. tables) for transfers
 - Makes the distinction between transfer format and database structure obvious
 - Relations are represented by position (containment in the XML hierarchy) instead of explicit links (foreign keys)
 - Object identification is implicit in the XML path instead of explicit by unique keys
- Distinction between exchange and transfer formats
- Repository access
 - by style sheets and XML tools
 - import to an XML database and use XML queries (e.g. XPath)
 - convert and import to a database system and use database tools

Result

Correct transfers !

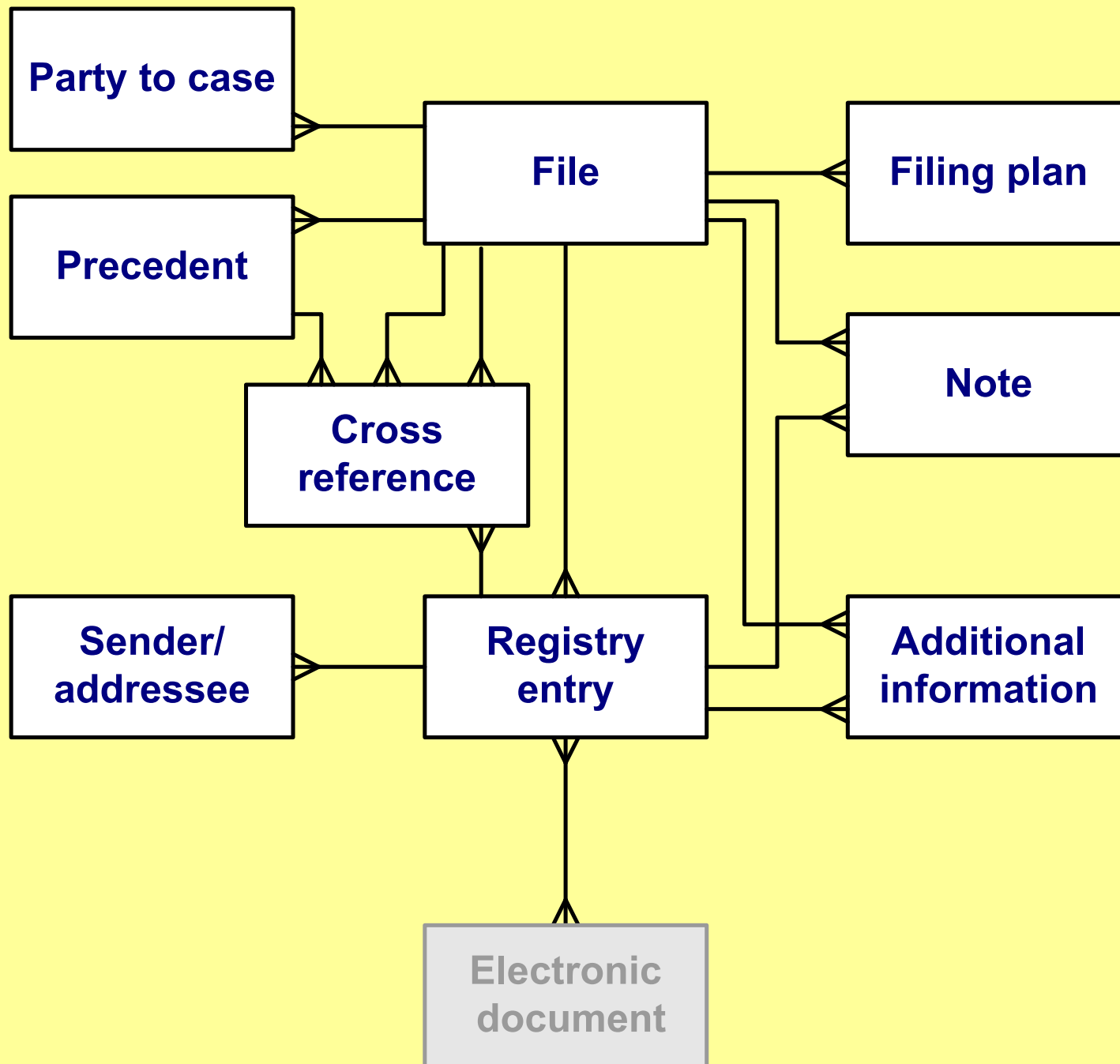
Thank you for your patience

Extras/details

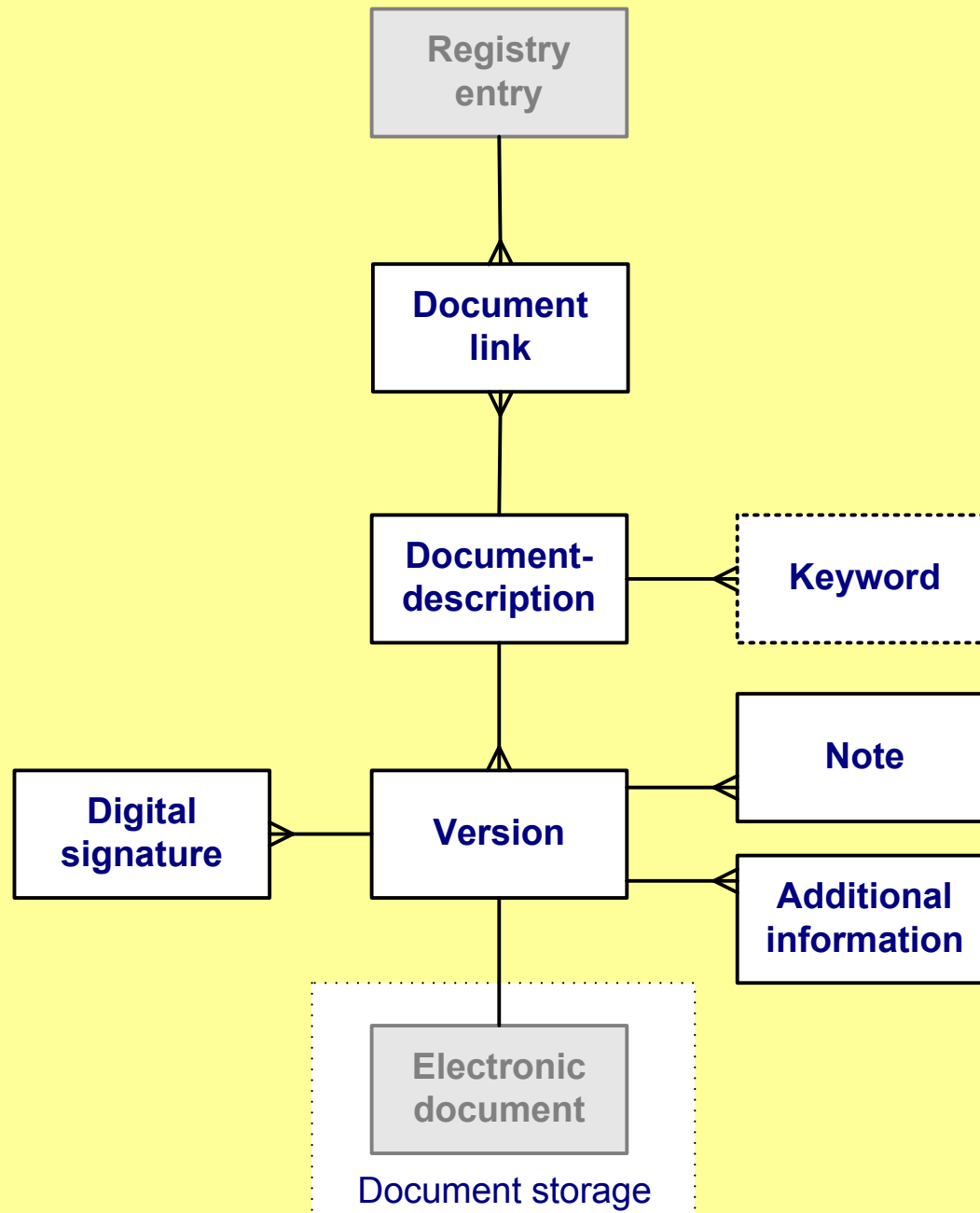
Noark v. 1-3 transfer format: The File table

Position	Field name	Length	Pic	Comments
1	Posttype	1	X	Fast verdi S
2-4	Grad	3	X(3)	
5-14	Saksnr.	10		
	<i>Årstall</i>	4	9(4)	jf. kommentar #1
	<i>Skråstrek</i>	1	X	Fast verdi /
	<i>Fortløpende nr.</i>	5	9(5)	Høyrejustert, nuller foran
15-22	Dato	8	9(8)	Form: ÅÅÅÅMMDD, jf. kommentar #1
23-24	Arkiv: 1. delfelt	2	X(2)	Angir delarkiv
25-40	Arkiv: 2. delfelt	16	X(16)	Arkivkode/Arkivbetegnelse
41-48	Blankt felt	8	X(8)	Fast verdi: Blank
49-52	Saksansv.: 1. delfelt	4	X(4)	Angir kontor
53-56	Saksansv.: 2. delfelt	4	X(4)	Saksansvarligs initialer
57-72	Presedens	16	X(16)	
73-74	Kassasjon	2	9(2)	Kassasjonskode, høyrejustert
75-90	Kryssreferanse	16	X(16)	
91-160	Sakstittel: 1. linje	70	X(70)	
161-230	Sakstittel: 2. linje	70	X(70)	
231-233	Antall dok	3	9(3)	Høyrejustert
234-241	Siste dok. (datofelt)	8	9(8)	Form: ÅÅÅÅMMDD, jf. kommentar #1
242	Sak avsluttet	1	X	
243-258	Sekundærkode	16	X(16)	Sekundærkodeangivelse
259-260	Blankt felt	8	X(8)	Fast verdi: Blank

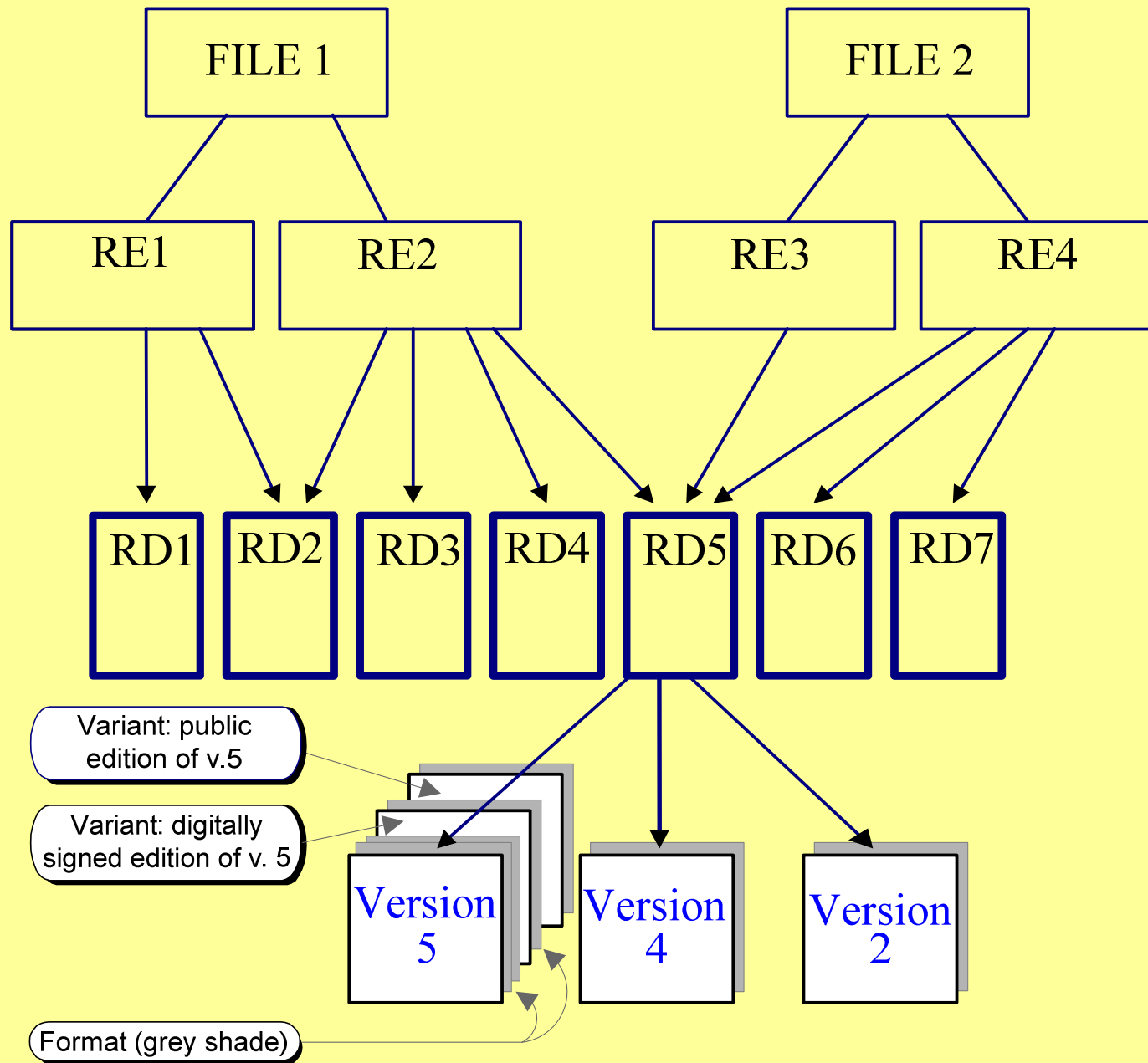
Noark-4 data model



Noark-4: Electronic document management (1)



Noark-4: Electronic document management (2)



Files

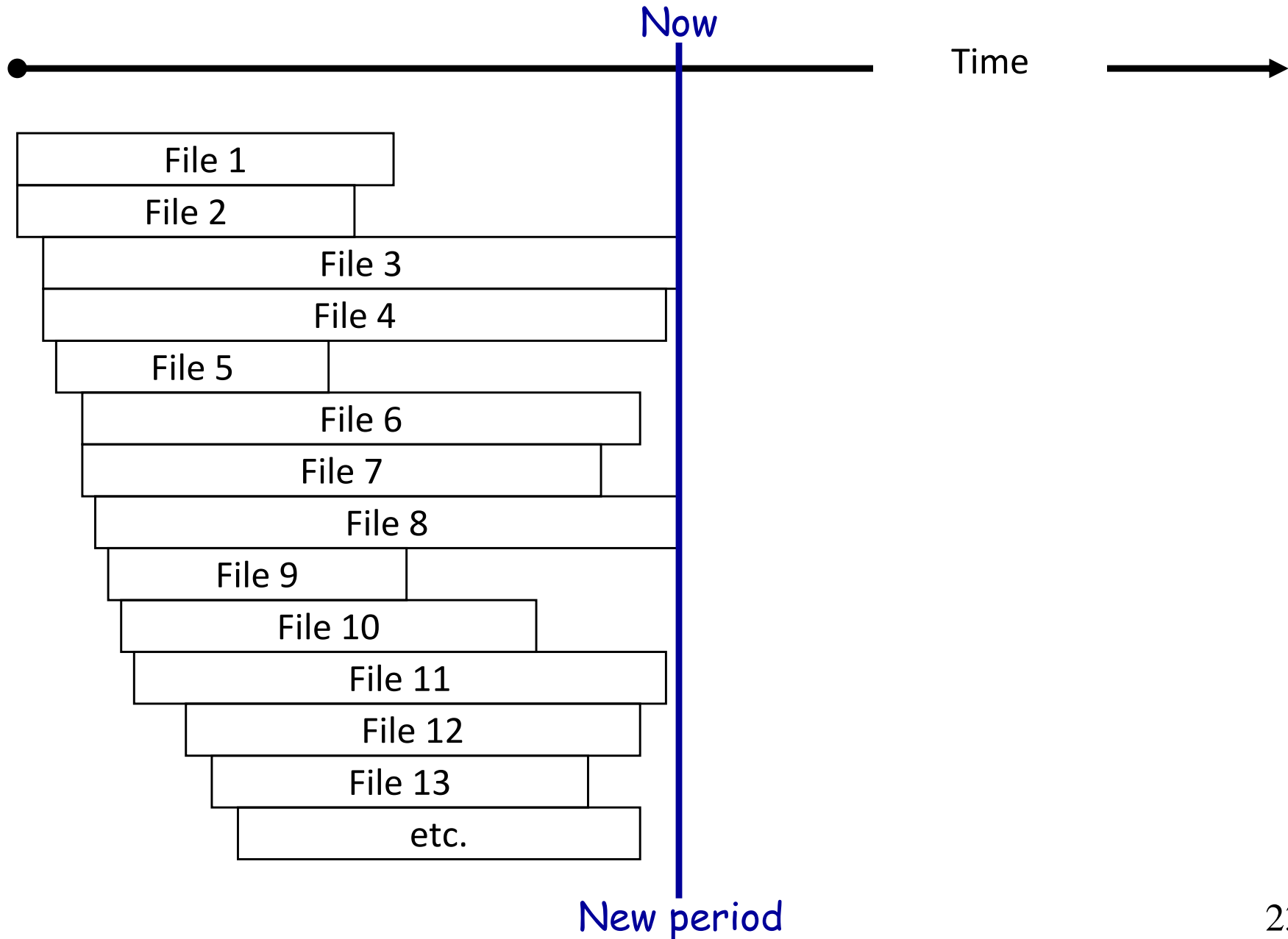
Registry entries

Document links

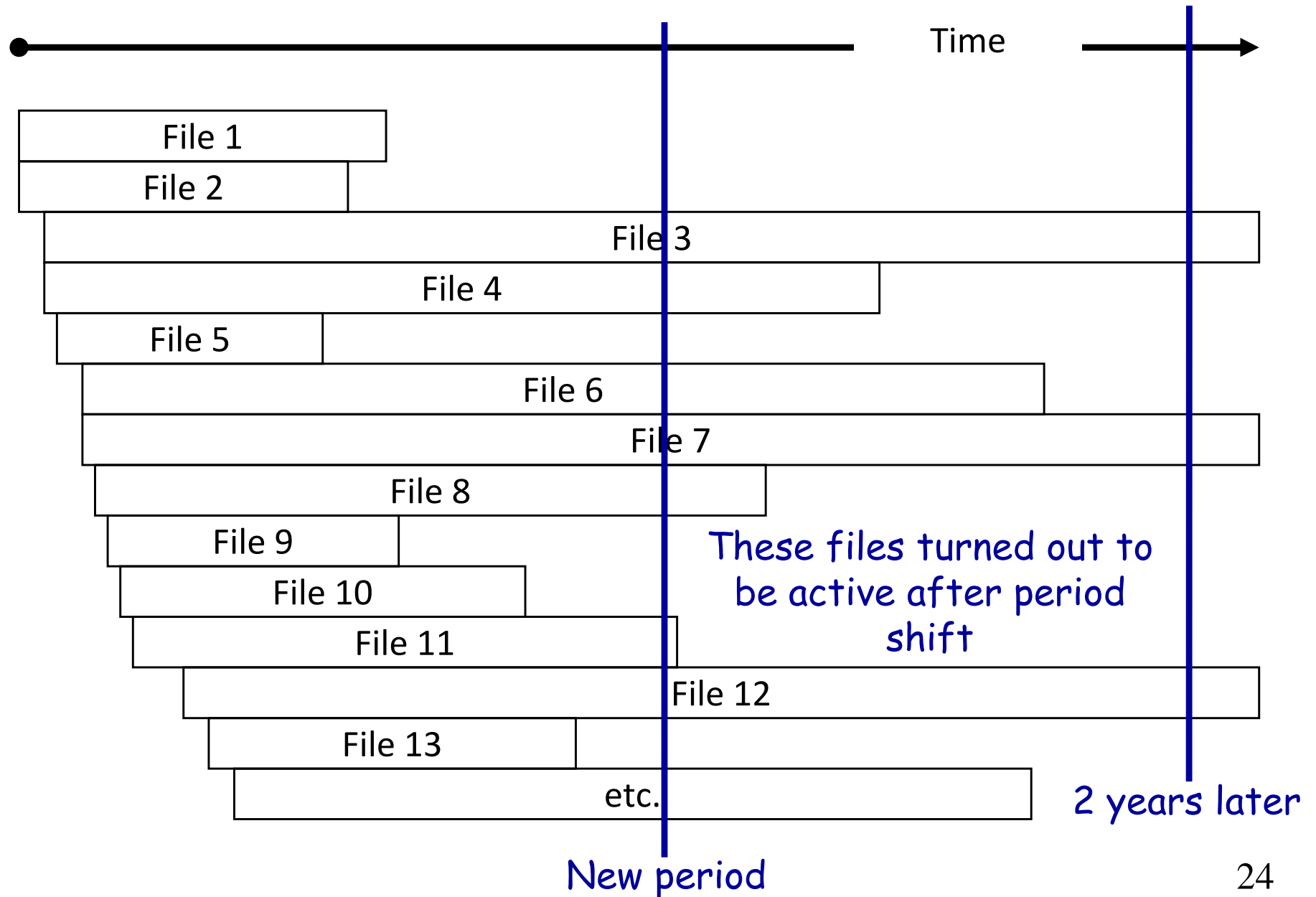
Record documents

Versions, variants (e.g. redactions) and formats

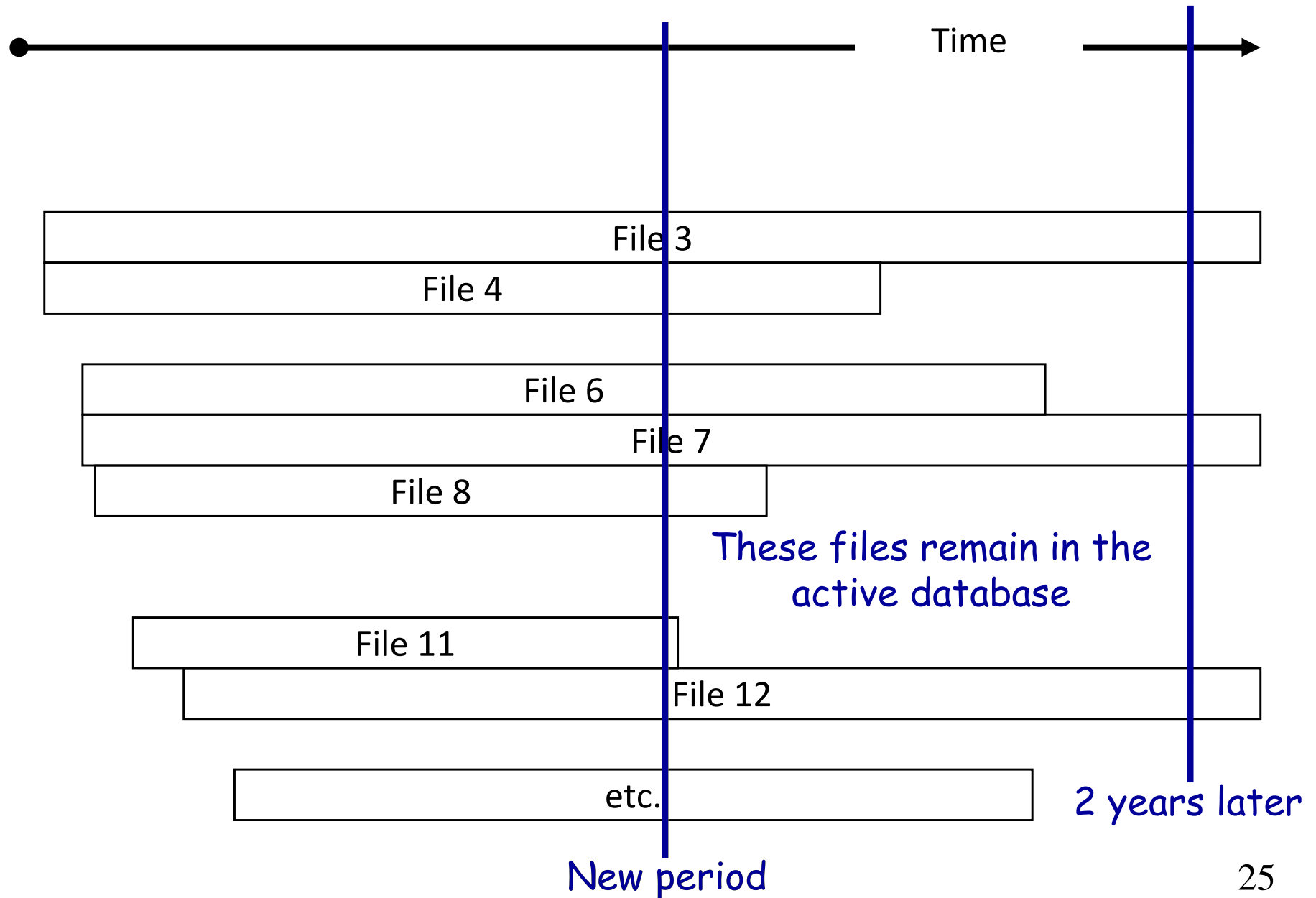
Archival periods in Noark systems (1)



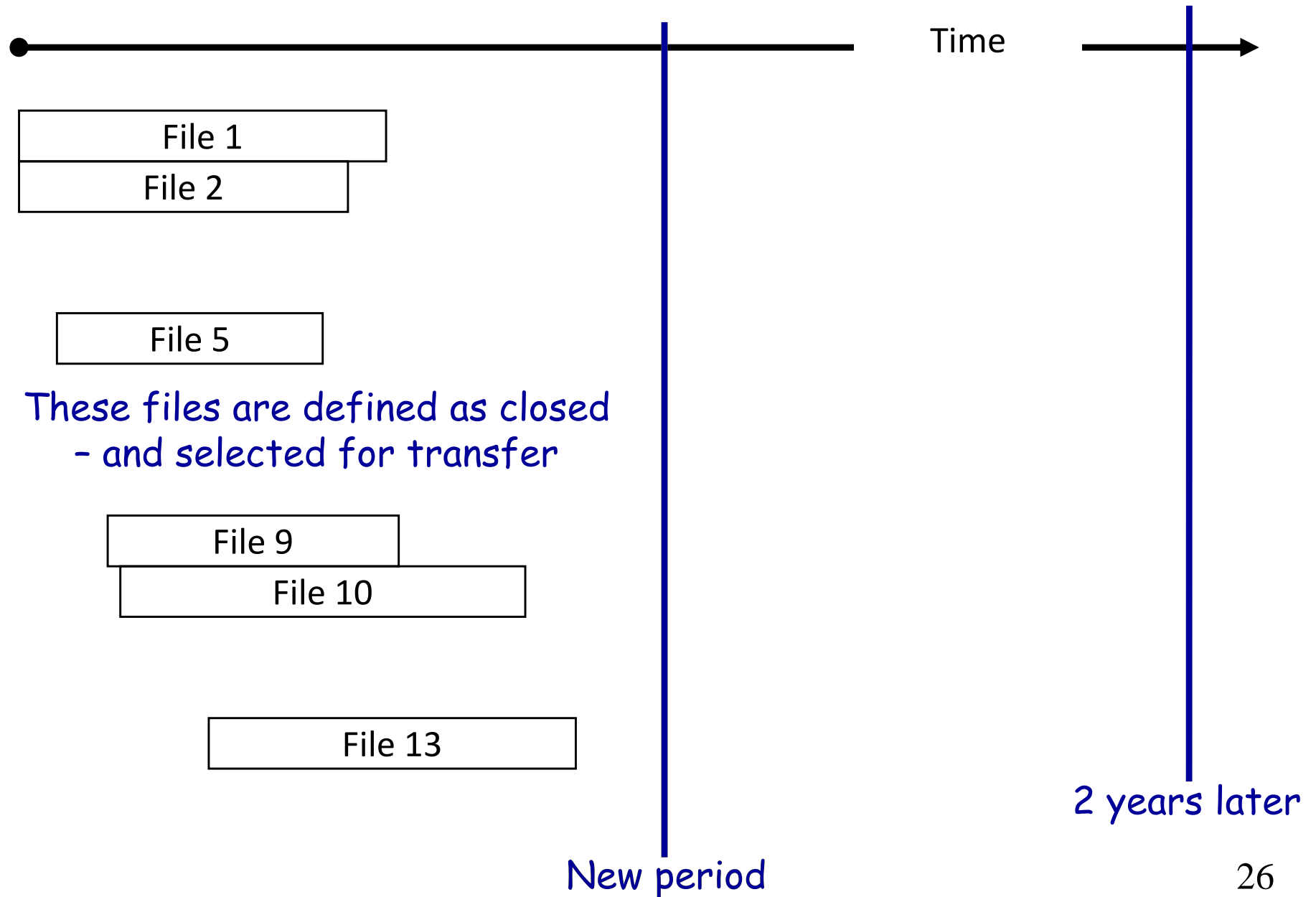
Archival periods in Noark systems (2)



Archival periods in Noark systems (3)



Archival periods in Noark systems (4)



To what extent can the Noark 5 transfer model be generalized?

- What do we want to preserve?
 - Data or archival records?
 - Systems or archival records?
- Is preservation dependent on the prolonged life of (regenerated) systems?
 - For reasons of integrity or authenticity?
 - Continued "system control" may be maintained by other means, for instance by the use of checksums
- Is the Noark 5 transfer model applicable outside the domain of recordkeeping systems?
 - Urgent need for simplification in response to the increasing complexity of systems
 - Necessary to focus on the distinction between logical information content and technical structure

Example: Business system of today on the logical vs. the technical implementation level

The (logical) presentation level selects data from 3 different databases.

What is the "system"? How do we define preservation?

The diagram illustrates the relationship between a logical presentation level and a technical implementation level. The logical presentation level is represented by a grid of boxes on the left, with arrows pointing to specific data points. The technical implementation level is represented by a screenshot of the ModuLink SaksArkiv software interface on the right, which displays a document schema for a specific case.

ModuLink SaksArkiv - Dokument skjema - 99/01098-005

Saksnr	1999 01098	Sdato	15.02.99	Saksansv	EARK TSI	
Arkiv	053	Edb - maskinvare og programvare				
Tittel						
Utviklingsprosjekt: Verktøy for å teste avleverte edb-arkiver m.m						
Ant dok	5	Siste dok	19.05.99	Obs		
Doknr	005	Jdato	19.05.99	Doktype	U	
Mottaker		Datert	03.05.99	Saksb	EARK TSI ?	
AVENIR ASA		Avskriver doknr				
Beskr						
Utviklingsprosjekt: Kontroll og konvertering av avleverte datafiler						
Tillegg						
arkivert						
Grad		Uoff		Avgrad		
Jenh	RA	Vedlegg		Forfall		
Lnr	005866 1999	Kode		Utlånt		
					Fulltekst	<input type="checkbox"/>