Policy Practice and Planning for a sustainable digital future

Workshop: Developing a National Digital Preservation Infrastructure

November 18th 2014, London

Marcel Ras & Joost van der Nat

Dutch National Coalition Digital Preservation

Agenda

NCDD
NATIONALE COALITIE DIGITALE DUURZAAMHEID

1. Introduction
2. NCDD

RESEARCH

3. Goals
4. Analytical framework
5. Matrix
6. Scenarios

WORKSHOP

7. Briefing
8. Work !
9. Plenary feedback

Building a future for our digital memory, together

NCDD
NATIONALE COALITIE DIGITALE DUURZAAMHEID

Netherlands Coalition for Digital Preservation

To establish an organisational and technical infrastructure for long-term access to all digital objects which are of crucial importance for science, culture and society

Why ?

NCDD
NATIONALE COALITIE DIGITALE DUURZAAMHEID

- The overall purpose why to foster cooperation is to increase both **Effectiveness** and **Efficiency**:
  - the ultimate goal is to enable all digital archiving organisations in The Netherlands to make and keep their digital objects accessible (effectiveness),
  - and to do this at costs that are as low as possible. This can be achieved by cooperating; economies of scale (efficiency).
- The assumption is that the more is cooperated, the bigger the delta is in increased Effectiveness and Efficiency.
- It is important to make a distinction between *administrative* responsibilities and *operational* responsibilities (that can be outsourced).
- The starting point for NCDD is: *share what is possible, keep separate what must be separate*.

Netherlands Coalition for Digital Preservation

NCDD
NATIONALE COALITIE DIGITALE DUURZAAMHEID

Established in 2008

Cross-domain national coalition covering the entire public sector

5 Members, representing a domain:

- The National Archive (NA)
- The National Library of The Netherlands (KB)
- Data Archiving and Networked Services (DANS)
- Dutch Institute for Sound and Vision (Beeld & Geluid)
- The Netherlands Cultural Heritage Coalition for Digital Preservation (CCDD)

The NCDD is

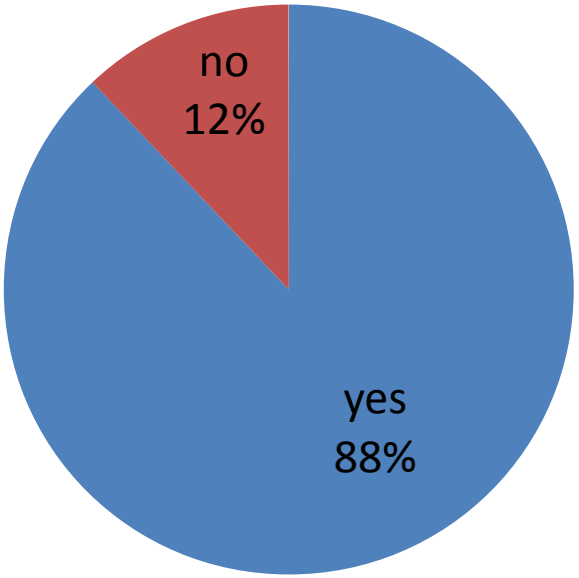
NCDD
NATIONALE COALITIE DIGITALE DUURZAAMHEID

Platform	Infrastructure	Bridge	Critical think tank
• Advocacy • Knowledge exchange • Knowledge development • Education & training	• Project coördinator • National infrastructure • Building Blocks	• International developments • International collaboration • Europe	• Policy making

What do we have?

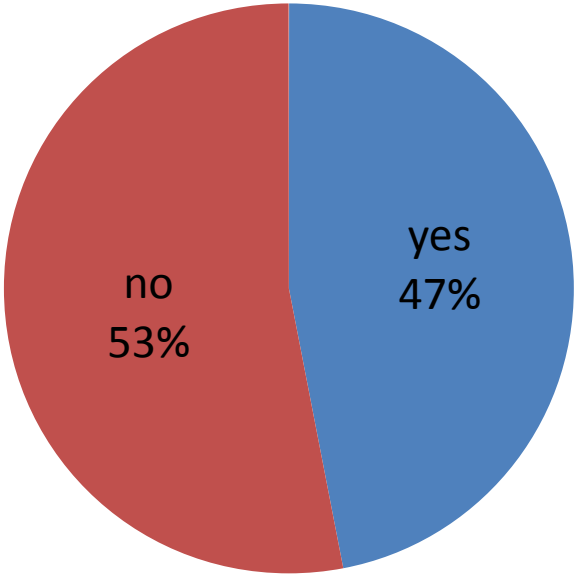
NCDD
NATIONALE COALITIE DIGITALE DUURZAAMHEID

Institution digitizes or
has digital collection (n = 141)



(European average: 87%)

Institution collects
born digital heritage (n = 98)



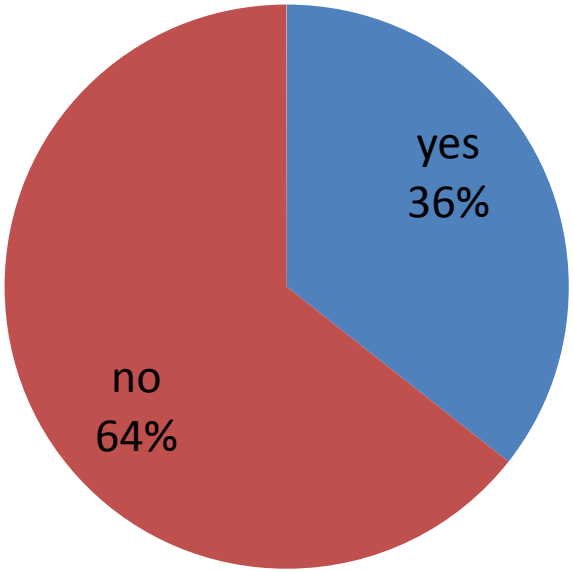
(European average: 53%)

ENUMERATE

What do we have not (yet)?

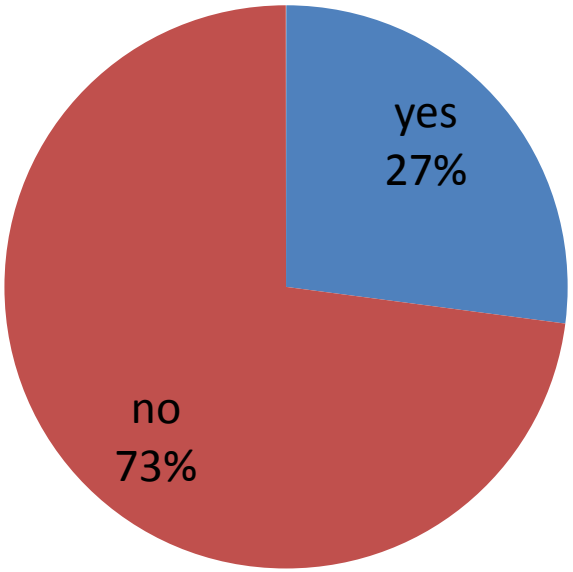
NCDD
NATIONALE COALITIE DIGITALE DUURZAAMHEID

Is there a policy for digital preservation / long term access?
(n = 101)



(European average: 26%)

Involved in an infrastructure for digital preservation?
(n = 96)



(European average: 30%)

ENUMERATE

National policy

NCDD
NATIONALE COALITIE DIGITALE DUURZAAMHEID

Towards a national approach:

- Connect existing infrastructures
- Responsibilities and roles for large knowledge institutions
- Three layer model:
  - Users, references, data (i.e. “visible”, “usable”, “preservable”)
- Roadmap: policy driven
- Cross-domain collaborative programme

Preservable:

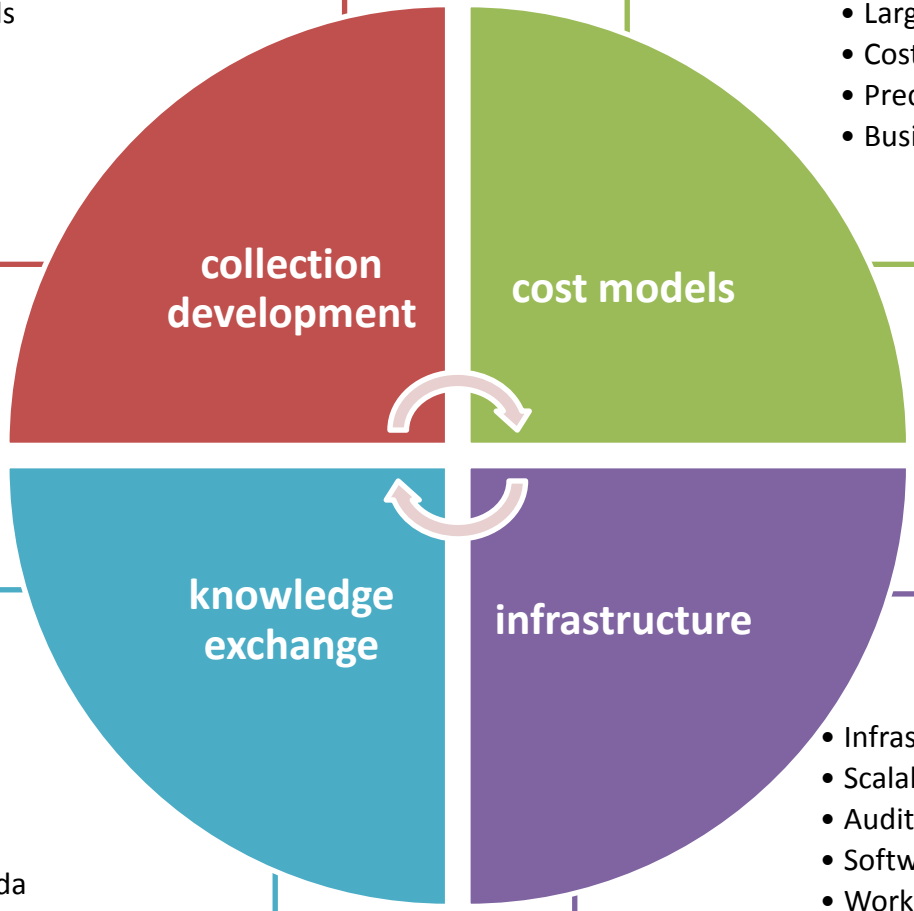
- Building a national infrastructure through cross-domain collaboration
- Collaboration = more effective and more efficient
- Three theme’s:
  - Scalability of the archives and infrastructure
  - Cost effectiveness
  - Roles and responsibilities in collection development

Building blocks in projects

NCDD
NATIONALE COALITIE DIGITALE DUURZAAMHEID

- Collection development policies
- Digital Collection Netherlands
- Roles & Responsibilities
- Automated QA
- Scalability
- Enhanced publications

- Cost management
- Large scale investment agenda
- Cost exchange
- Predictive models
- Business models



- Knowledge exchange
- Staff training
- Advocacy
- Communication
- Research & Innovation agenda

- Infrastructure AS IS and TO BE
- Scalable infrastructure
- Audit & certification
- Software archive
- Workflows
- Storage

Research national infrastructure

NCDD
NATIONALE COALITIE DIGITALE DUURZAAMHEID

Agenda

NCDD
NATIONALE COALITIE DIGITALE DUURZAAMHEID

1. Introduction
2. NCDD

RESEARCH

3. Goals
4. Analytical framework
5. Matrix
6. Scenarios

WORKSHOP

7. Briefing
8. Work !
9. Plenary feedback

Goals of the research

(start April – finish December 2014)

NCDD
NATIONALE COALITIE DIGITALE DUURZAAMHEID

1. Update of the picture of the “AS IS” situation of the present infrastructure for digital preservation (DPI) across the 5 domains
2. Present costs incurred by the domains for DPI
3. Estimate of growth (volume, complexity, diversity)
4. Description of the “TO BE” situation of the infrastructure for DP
5. A “FIT – GAP” analysis
6. Scenario’s for transition from “AS IS” to “TO BE”
7. Estimate of required DPI budgets (next 5 years, 5 years after)
8. “Prediction model” for estimating growth (volume and €)
9. Definition of next phase, amongst which the development & implementation of a (national ?) cost model....

State of affairs, done

NCDD
NATIONALE COALITIE DIGITALE DUURZAAMHEID

- Started early April 2014
- Desk research (400+ articles, books, websites)
- Developed the analytical framework
- Conducted 10 interviews with organisations
- 15^{+} interviews with experts, stakeholders (OCW !)
- Processed results
- Developed scenarios
- Developed diagrams
- Concept Table of Contents for Final Report

State of affairs, to be done

NCDD
NATIONALE COALITIE DIGITALE DUURZAAMHEID

- Finalise interviews:
  - Nationaal Museum van Wereldculturen
  - the CIO of the ministry of Education, Culture and Science (OCW)
  - vendors
- Expand the results of the research
- Elaborate on scenarios (18 attributes, e.g. Transition scenarios)
- Look into the feasibility of the financial project goals within this project (present, future, “predictability model”)
- Verification & validation of results
  (NCDD members, interviewees, stakeholders, OCW)

- Finalise the Report

Agenda

NCDD
NATIONALE COALITIE DIGITALE DUURZAAMHEID

1. Introduction
2. NCDD

RESEARCH

3. Goals
4. Analytical framework
5. Matrix
6. Scenarios

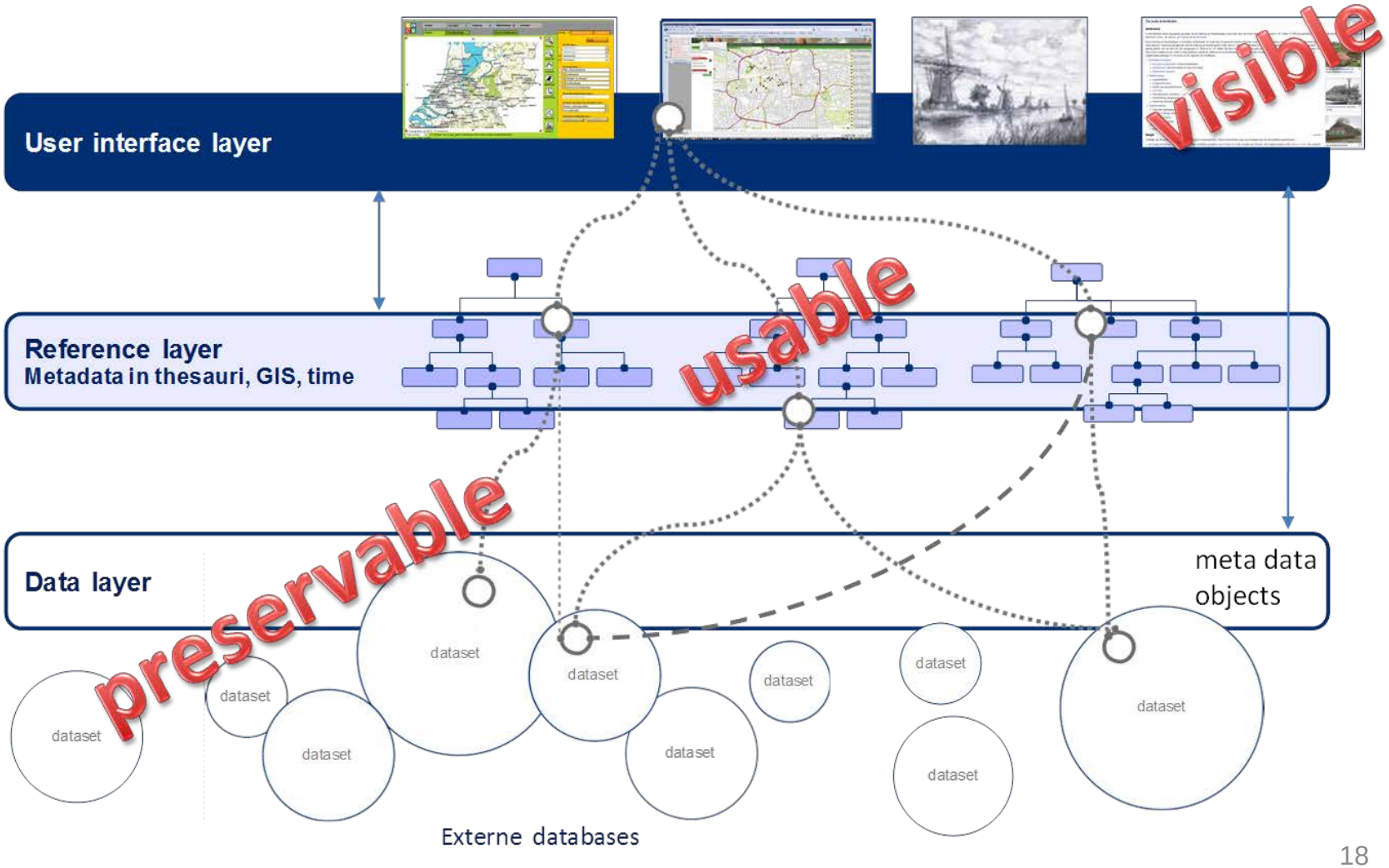
WORKSHOP

7. Briefing
8. Work !
9. Plenary feedback

Analytical framework, 3 layers

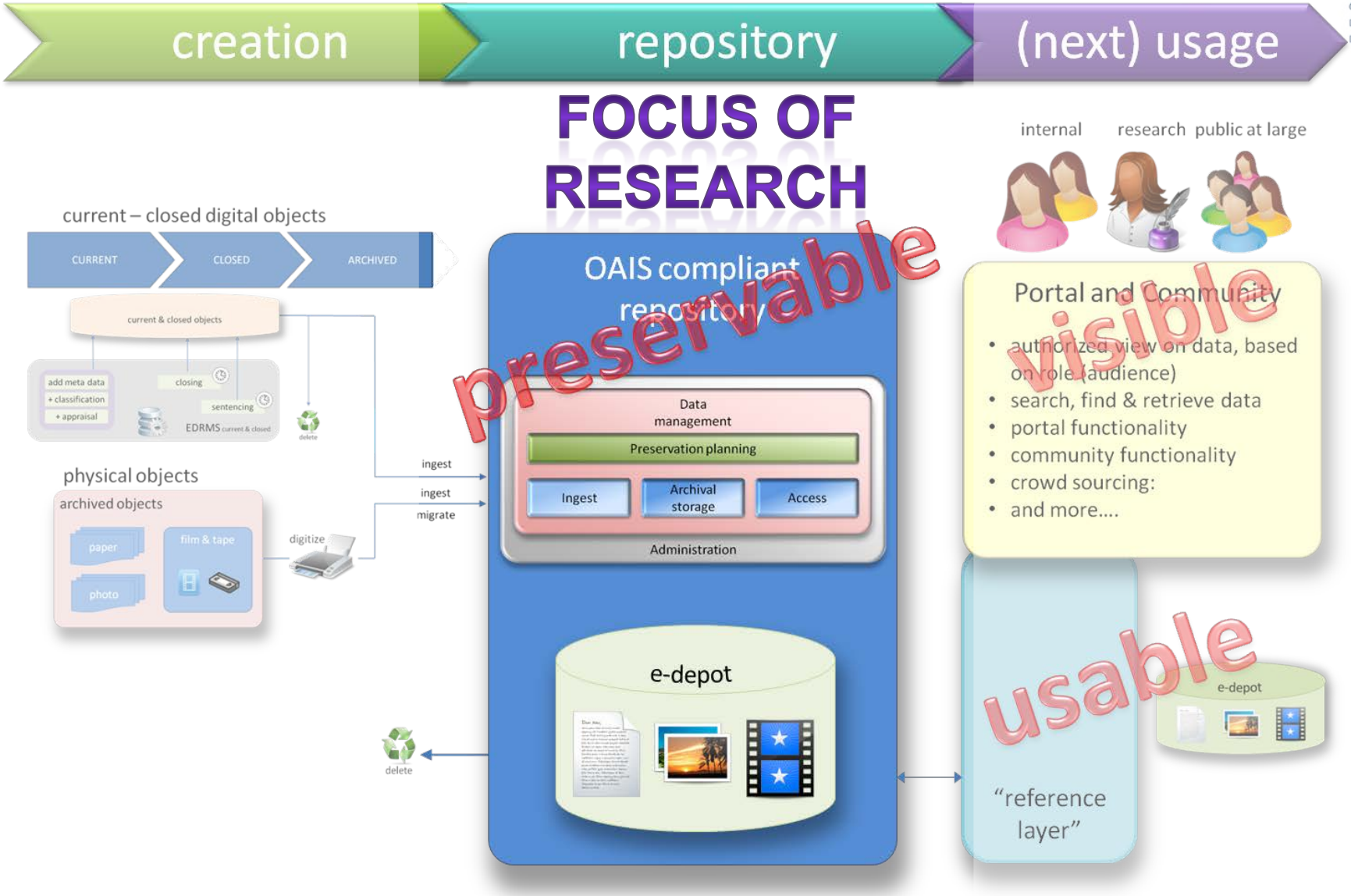
model of the Ministry of OCW

NCDD
NATIONALE COALITIE DIGITALE DUURZAAMHEID

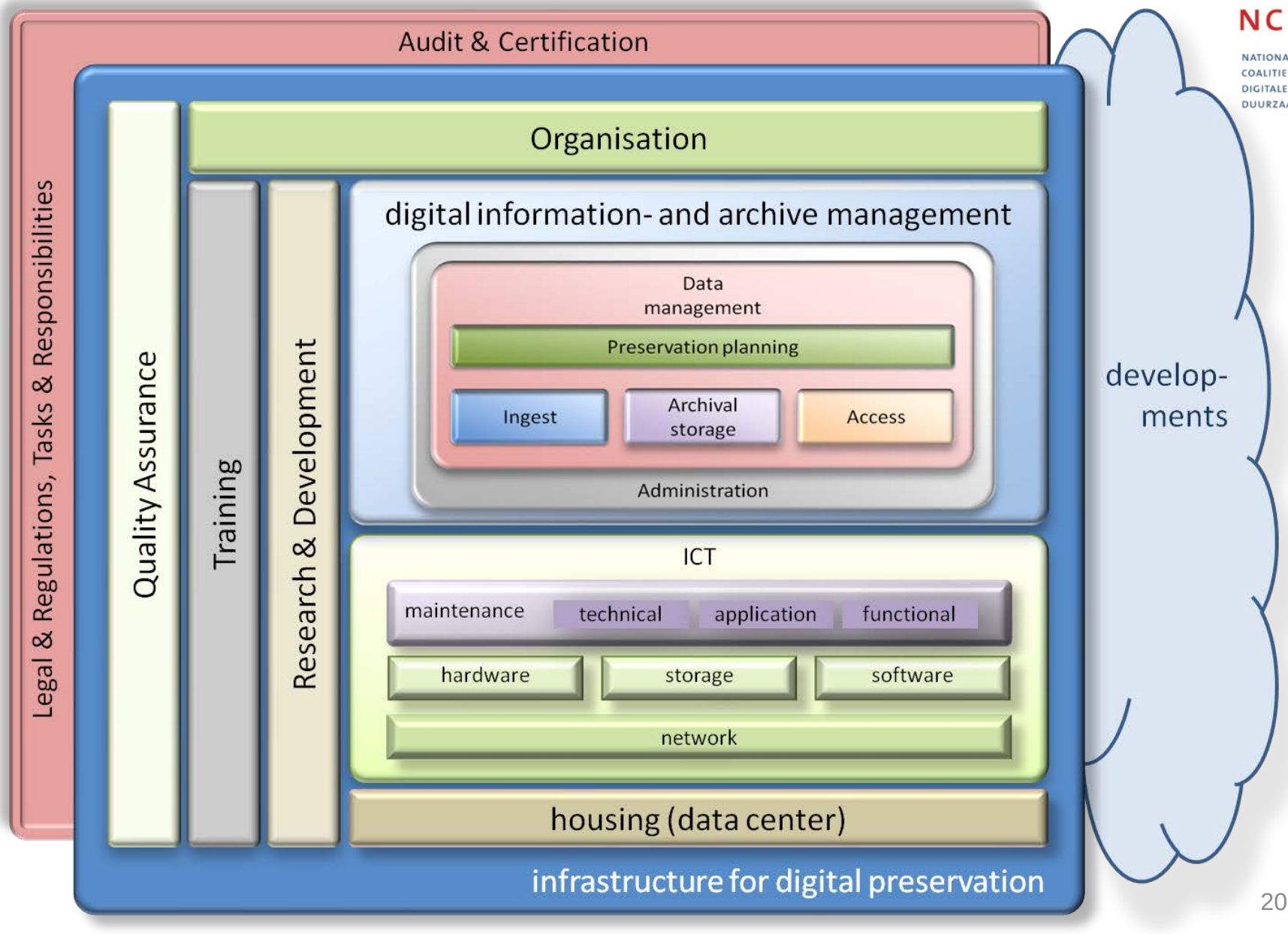


Analytical framework, 3 columns

model of NCDD



Analytical framework, “infrastructure”



NCDD
NATIONALE COALITIE DIGITALE DUURZAAMHEID

Analytical framework,

APARSEN WP 21

	High-level service	Ingest	Preservation Planning	Data Management	Archival Storage	Administration	Access
1	Characterisation of SIPs	●					
2	Quality assurance of SIPs	●					
3	Policy-based assessment of SIPs	●					
4	Acquisition and maintenance of represent. info	●	●			❍	
4.1	Automated metadata creation/maint	●				●	
4.1.1	Metadata migration					●	
5	Environment monitoring (preservation watch)		●				
5.1	Knowledge model comparison		●				
6	Preservation plan formulation		●				
6.1	Obsolescence substitution		●				
6.2	Dependency management		●				
7	Authenticity evidence management		●				
8	Appraisal of collections		●	❍	❍	❍	
9	DRM clearinghouse		●				●
10	Brokerage between repositories		●				
11	Long-term archiving	❍		❍	●	❍	❍
11.1	Integrity checking				●		
11.2	Cloud storage for preservation				●		
12	Preservation policy construction					●	
13	Analysis of authenticity management policies					●	
14	Format transformation	●				●	
15	Finding aids						●
15.1	Federated search						●
16	PID resolver						●
17	Emulation facilities		❍				●
18	Full repository service	●	●	●	●	●	●
19	Audit and certification of repositories	●	●	●	●	●	●

● service is a key contributor to the corresponding OAIS functional entity

❍ indicates possible or marginal relevance

NCDD
NATIONALE COALITIE DIGITALE DUURZAAMHEID

Analytical framework, overall

NCDD
NATIONALE COALITIE DIGITALE DUURZAAMHEID

CREATION		DIGITAL REPOSITORY					NEXT USAGE
		Primary Process of Repository			Support	Secondary	
Creation	Handover	Ingest	Archival Storage	Access	Primary Process	Support	Handover NEXT USAGE
Create object		Characterisation of SIPs	Long-term archiving	DRM clearinghouse	**Preservation Planning**	Quality Control	by designated communities or other repositories
		Quality assurance of SIPs	Integrity checking	Finding aids	Acquisition and maintenance of represent. info	Training	
Appraisal of collections		Policy-based assessment of SIPs	Cloud storage for preserv	Federated search	Environment monitoring (preservation v	R & D	
Characterisation of SIPs		Acquisition and maintenance of represent. info		PID resolver	Knowledge model comparison	Audit & Certification	ingest
Quality assurance of SIPs		Automated metadata creation/maint		Emulation facilities	Preservation plan formulation		Archival Storage
Policy-based assessment of SIPs		Format transformation			Obsolescence substitution		Access
Acquisition and maintenance of represent. info					Dependency management		Support Primary Process
Automated metadata creation/maint					Authenticity evidence management		Secondary Support
Format transformation					Appraisal of collections		*(each of which has services as mentioned to the left...)*
Brokerage between repositories					DRM clearinghouse		
Authenticity evidence management		REMARKS:			Brokerage between repositories		
Appraisal of collections		NOT all 27 APARSEN services are mapped here					
DRM clearinghouse		- cells this colour appear in more than 1 column...			**Data management**		
					Administration		
					Automated metadata creation/maint		
					Metadata migration		
					Preservation policy construction		
					Analysis of authenticity management policies		
					Format transformation		
IT maint. technical application functional		IT maintenance technical application functional					IT maint. technical application functional
software storage hardware		software storage hardware					software storage hardware
Housing (data center)		Housing (data center)					Housing (data center)

Agenda

NCDD
NATIONALE COALITIE DIGITALE DUURZAAMHEID

1. Introduction
2. NCDD

RESEARCH

3. Goals
4. Analytical framework
5. Matrix
6. Scenarios

WORKSHOP

7. Briefing
8. Work !
9. Plenary feedback

Matrix, preliminary results

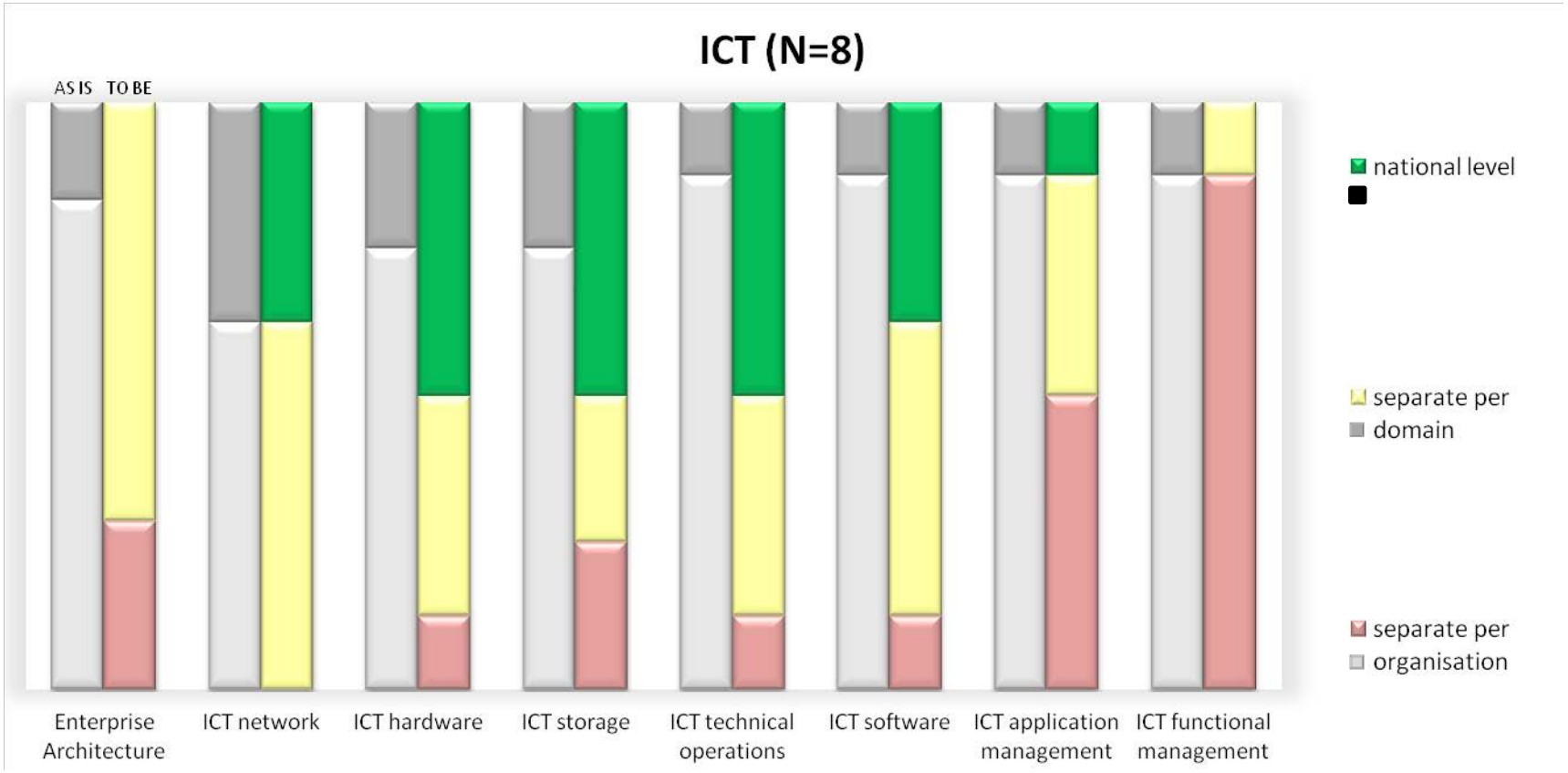
versie 5, 19-09-2014

Elements		OVERALL TO BE and AS IS	separate per organisation	separate per domain	national level
Systems	Enterprise Architecture		①③-⑤-⑦⑨❸-❺	⑧❼❾❶❻❽	
	ICT network		③⑥⑦⑧⑨	①④⑤❶❹❺❻❾	❸❼❽
	ICT hardware		①③⑤⑥⑦⑨❺	④⑧❹❻❾	❶❸❼❽
	ICT storage		①③⑤⑥⑦⑨❺❾	④⑧❹❻	❶❸❼❽
	ICT technical operations		①③④⑤⑥⑦⑨❺	⑧❹❻❾	❶❸❼❽
	ICT software		①③④⑤⑥⑦⑨❺	⑧❹❻❽❾	❶❸❼
	ICT application management		①③④⑤⑥⑦⑨❶❸check❹❺	⑧❻❽❾	❼
	ICT functional management		①③④⑤⑥⑦⑨❶❸check❹❺❻❼❾	⑧❽	
Pre Primary Process	Meta data	obtain	①③-⑤⑥⑦⑧⑨❶-❺❻❼❾	❸❽	
		conversion	①③-⑤⑥⑦⑧⑨❶❸-❺❻❼❾	❽	
		enrichment	①③-⑤⑥⑦⑧⑨❶-❺❻❼❾	❸❽	
Primary process	Ingest		①❶		
		SIP-check	①③④⑤⑥⑦⑧⑨❶❸❹❺❻❼❾	❽	
		meta data obtain / maintenance	①③④⑤⑥⑦⑧⑨❶❸❺❻❼❾	❹❽	
		meta data automated	--④⑤⑥⑦⑧⑨-❸❺❻❾	❹❼❽	
		format transformation	①-⑤⑥⑦⑧⑨❶❺❻❼❾	④❹❽	❶❸
	Archival Storage		①		❶
		appraisal	①③-⑥⑦⑧⑨-❻❼❾	⑤❽	❺❶❸
		long term storage (Preservation as a service)	①-⑥⑦⑧⑨❾	④❹❼❽	⑤❶❸❺❻
		integrity check	①-④⑥⑦⑧⑨❾	❹❼❽	⑤❶❸❺❻
	Access		①❶		
		DRM clearinghouse	①③⑤⑥⑦⑧⑨❶❺❻❼❾	④❸❹❽	
		finding aids	①③⑤⑥⑦⑧⑨❶❻❼	❸check❽❾	④❺❹
		federated search	①--⑥⑦⑧⑨❶	❼❾	⑤❺❸check❹❻❽
		PID resolver	①③⑤--⑧	④	⑨❶❸❹❺❻❼❽❾
		emulation facilities	------⑧--	❻	❸❹-❼❽❾
Support of Primair Proces	Preservation Planning		-	❶	
		preservation watch (Format Registry)	-③④⑤⑥⑦⑨	❶❻	⑧❸❹❺❼❽❾
		preservation plan formulation	-③④⑤⑥-⑨❸❺	⑦❾❶❹❼❽	❻
		dependency management	-③-⑤⑥--⑨❺	❾❶❹❼❽	❻❸
		authenticity evidence management	-③-⑤⑥⑦⑧⑨❸	❾❶❹❼❽	❻❺
		exchange between repositories	-③⑥⑧⑨	④❶❸❹❾	⑦⑤❺❻❼❽
	Data Management		①③⑤⑥⑦⑧⑨❶❸❺❻❼❾	④❹❽	
	Administration		①⑤⑧❺	❶❽	
		preservation policy construction	-③④⑤⑥⑨❹❺❾	⑦⑧❸❶❼❽	❻
	Preservation tools		-③⑤⑥⑦⑧⑨	❶	④❸❹❺❻❼❽❾
Secondary support	Quality Assurance		①③④⑤⑥⑧⑨❻	⑦❺❼❾❶❸❹❽	
	Training		①③-⑦⑧⑨	④❺❶❹❼❽	⑤❸❻❾
	Research and development		①⑥⑧⑨	⑦④❻❺❶❹❽	⑤③❸❾❼
Audit & Certification			①③check-⑦⑧⑨	⑤④❶❼	❸check❹❺❻❽❾

LEGENDA

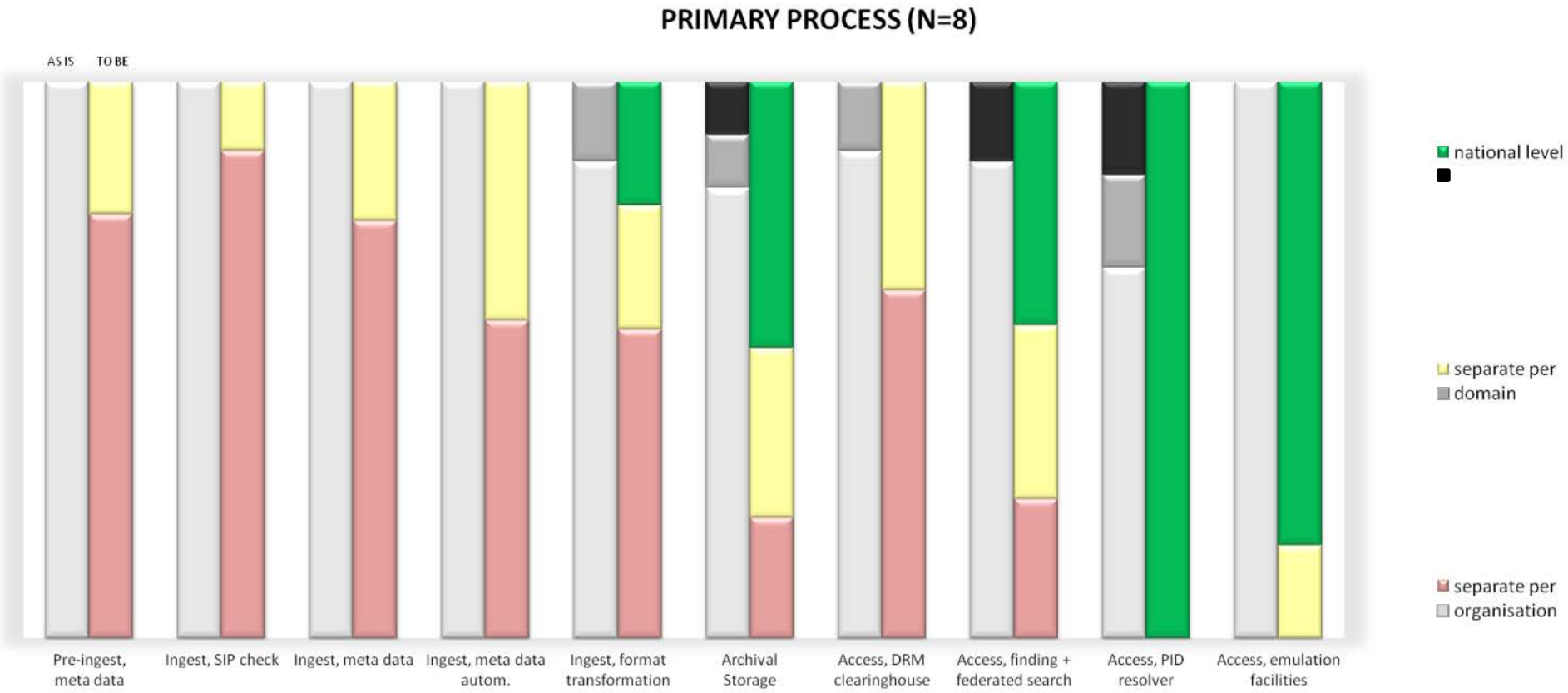
1 Rijksmuseum
~~2 Rijksdienst voor het Cultureel Erfgoed (RCE)~~
3 Koninklijke Bibliotheek
4 DANS
5 Instituut voor Beeld & Geluid
6 Het Nieuwe Instituut
7 Stadsarchief ASD
8 Nationaal Archief
9 IISG
10 LI-MA
11 Rijksmuseum Volkenkunde

Preliminary results, matrix
ICT

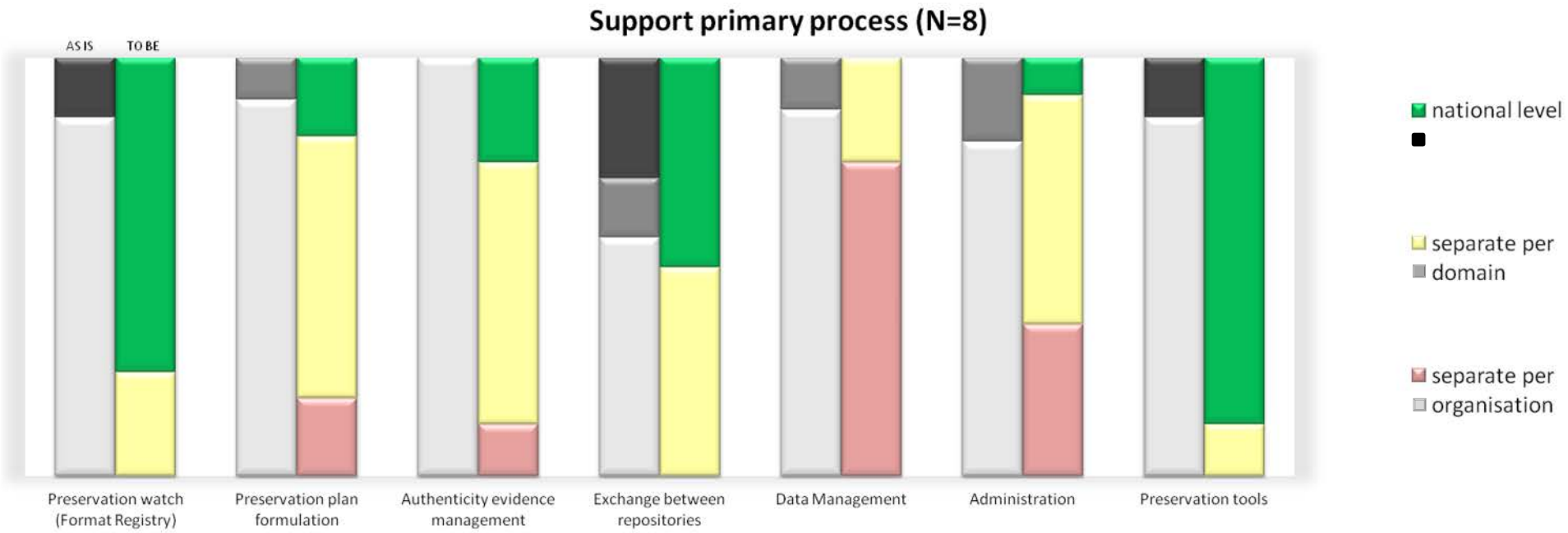


Preliminary results, matrix
Primary Process

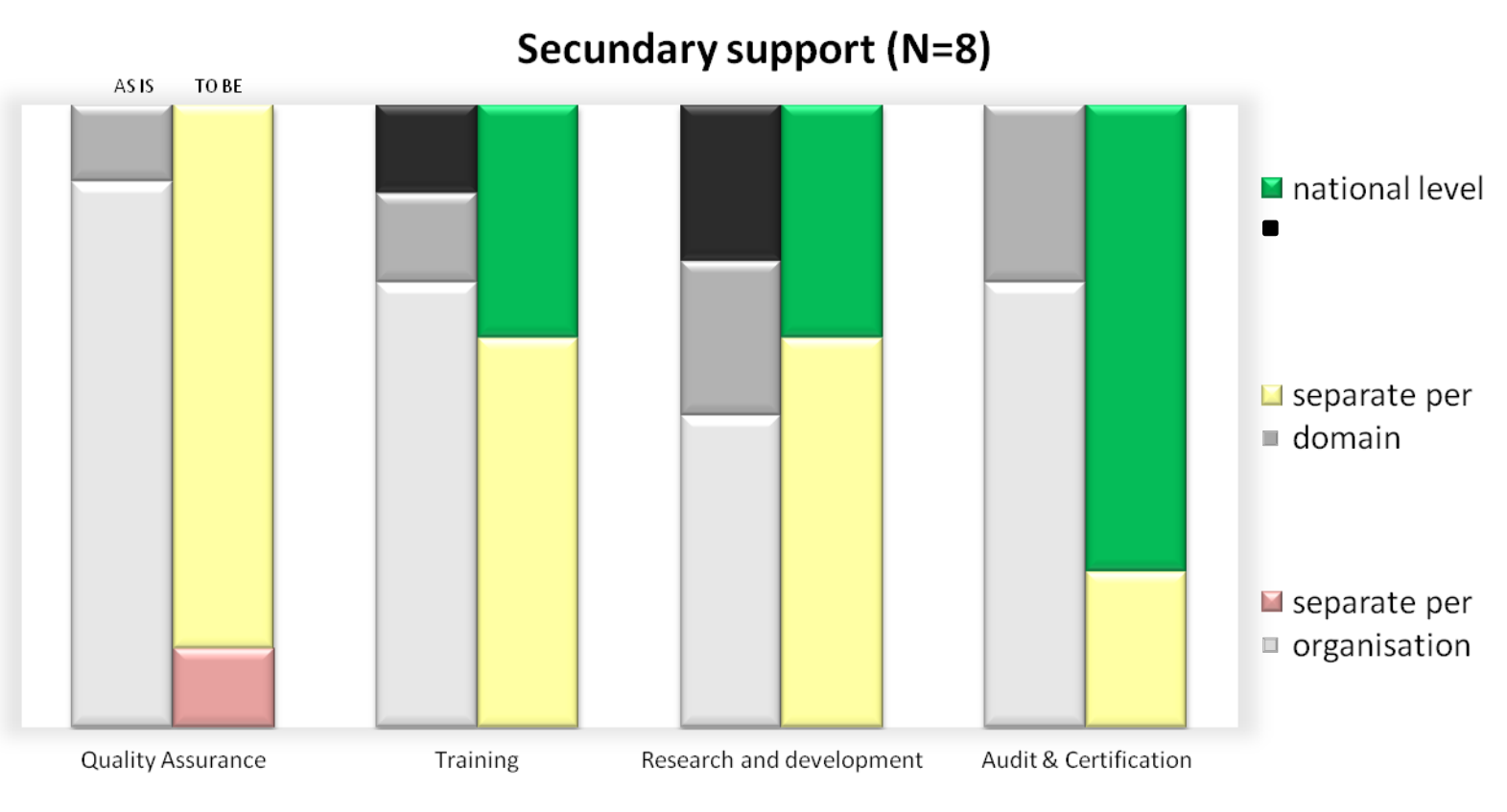
NCDD
NATIONALE COALITIE DIGITALE DUURZAAMHEID



Preliminary results, matrix
Support Primary Process



Preliminary results, matrix
Secundary support



Agenda

NCDD
NATIONALE COALITIE DIGITALE DUURZAAMHEID

1. Introduction
2. NCDD

RESEARCH

3. Goals
4. Analytical framework
5. Matrix
6. Scenarios

WORKSHOP

7. Briefing
8. Work !
9. Plenary feedback

Scenarios, goal

NCDD
NATIONALE COALITIE DIGITALE DUURZAAMHEID

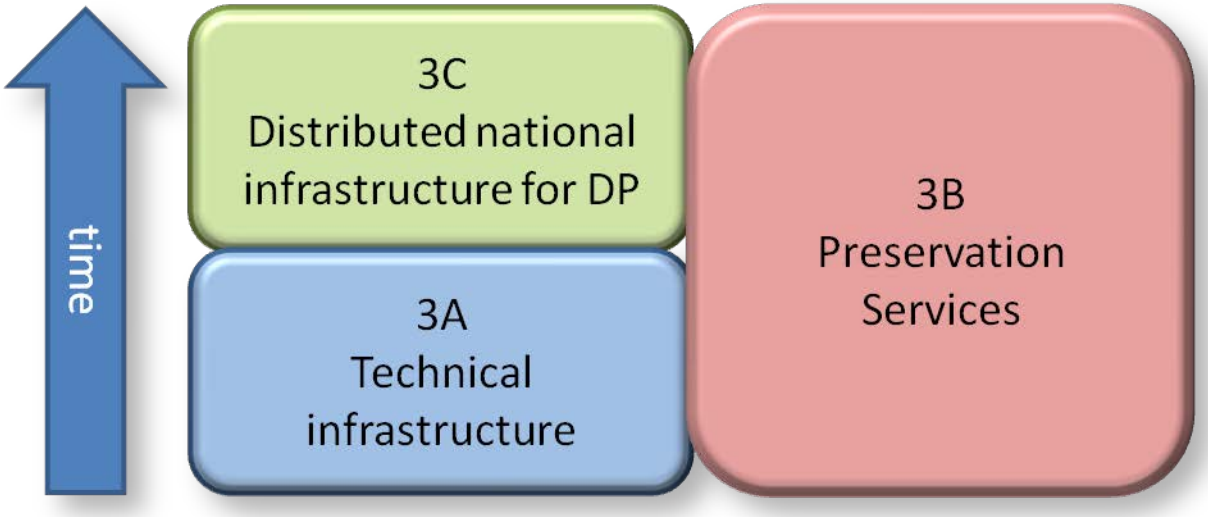
- The overall purpose why to foster cooperation is to increase both **Effectiveness** and **Efficiency**:
  - the ultimate goal is to enable all digital archiving organisations in The Netherlands to make and keep their digital objects accessible (effectiveness),
  - and to do this at costs that are as low as possible. This can be achieved by cooperating; economies of scale (efficiency).
- The assumption is that the more is cooperated, the bigger the delta is in increased Effectiveness and Efficiency.
- It is important to make a distinction between *administrative* responsibilities and *operational* responsibilities (that can be outsourced).
- The starting point for NCDD is: *share what is possible, keep separate what must be separate*.

Scenarios

NCDD
NATIONALE COALITIE DIGITALE DUURZAAMHEID

SCENARIO	
1	Doing nothing extra
2	Ad Hoc (extend present practice)
3	Shared Services
3 A	Technical infrastructure
3 B	Digital Preservation Services
3 C	Distributed National Infrastructure Digital Preservation
4	One NL e-Depot

Scenario 3, elements & time



- Technical infrastructure (3A) and Preservation Services (3B) run parallel
- Technical infrastructure (3A) is conditional for the Distributed national infrastructure (3C)
- First steps for 3A are already in progress: Rijksdata centres (all governmental data are to be stored in 4 data centres)

Scenarios, quick & dirty scan

based on “gut feeling”

NCDD
NATIONALE COALITIE DIGITALE DUURZAAMHEID

SCENARIO	effort required	feasibility	Δ effectivity	Δ efficiency
1 Doing nothing extra	○	●	○	○
2 Ad Hoc	◔	●	◔	◔
3 Shared Services				
3A. Technical infrastructure	◕	◕	◕	◕
3B. Digital Preservation Services	◑	◕	◕	◕
3C. Distributed National Infra-structure Digital Preservation	●	◑	●	●
4 One NL e-Depot	●	○	◑	◕

Scenarios, attributes

- Short description
- General purpose (why do it ?)

- Point of departure (axioms / postulates)
- Constraints (what must be in place to make it happen)

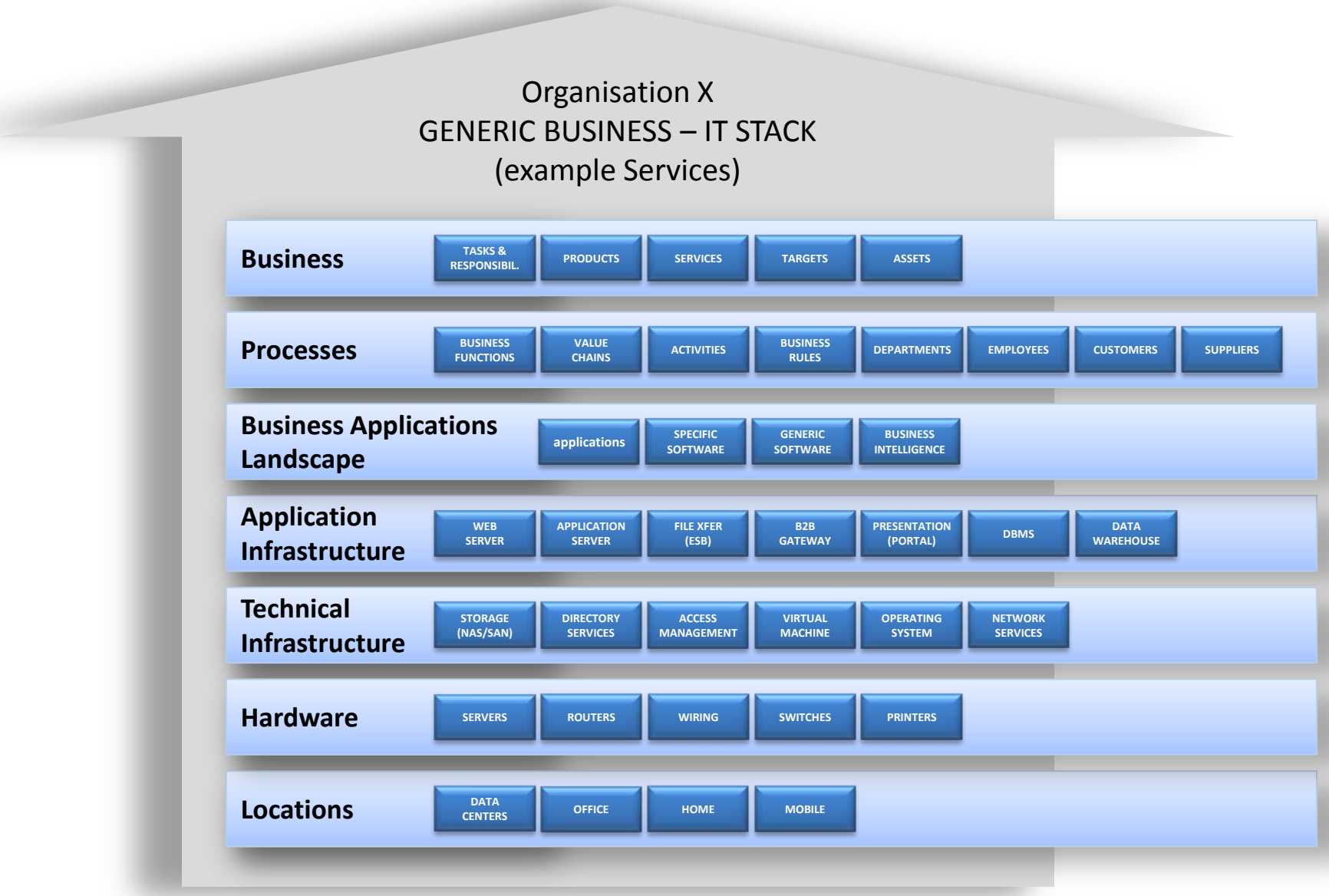
- Drivers (how)
- Enablers (what)

- Barriers (why not)
- Risks of scenario
- Pro's of scenario
- Con's of scenario

- Transition
  - instruments
  - the ways to walk
  - time path
- Risks of transition

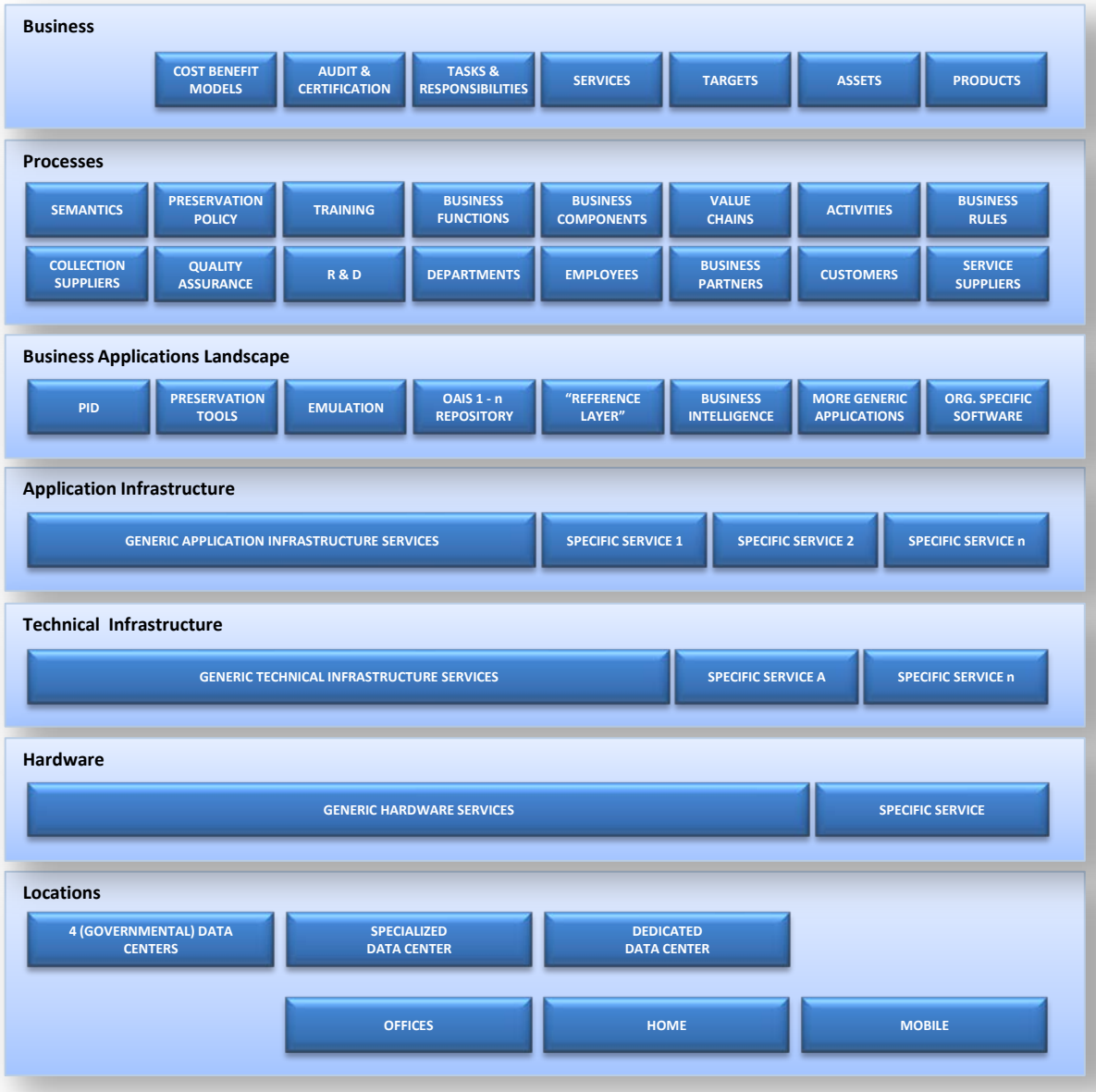
Scenario 3, building blocks general

NCDD
NATIONALE COALITIE DIGITALE DUURZAAMHEID



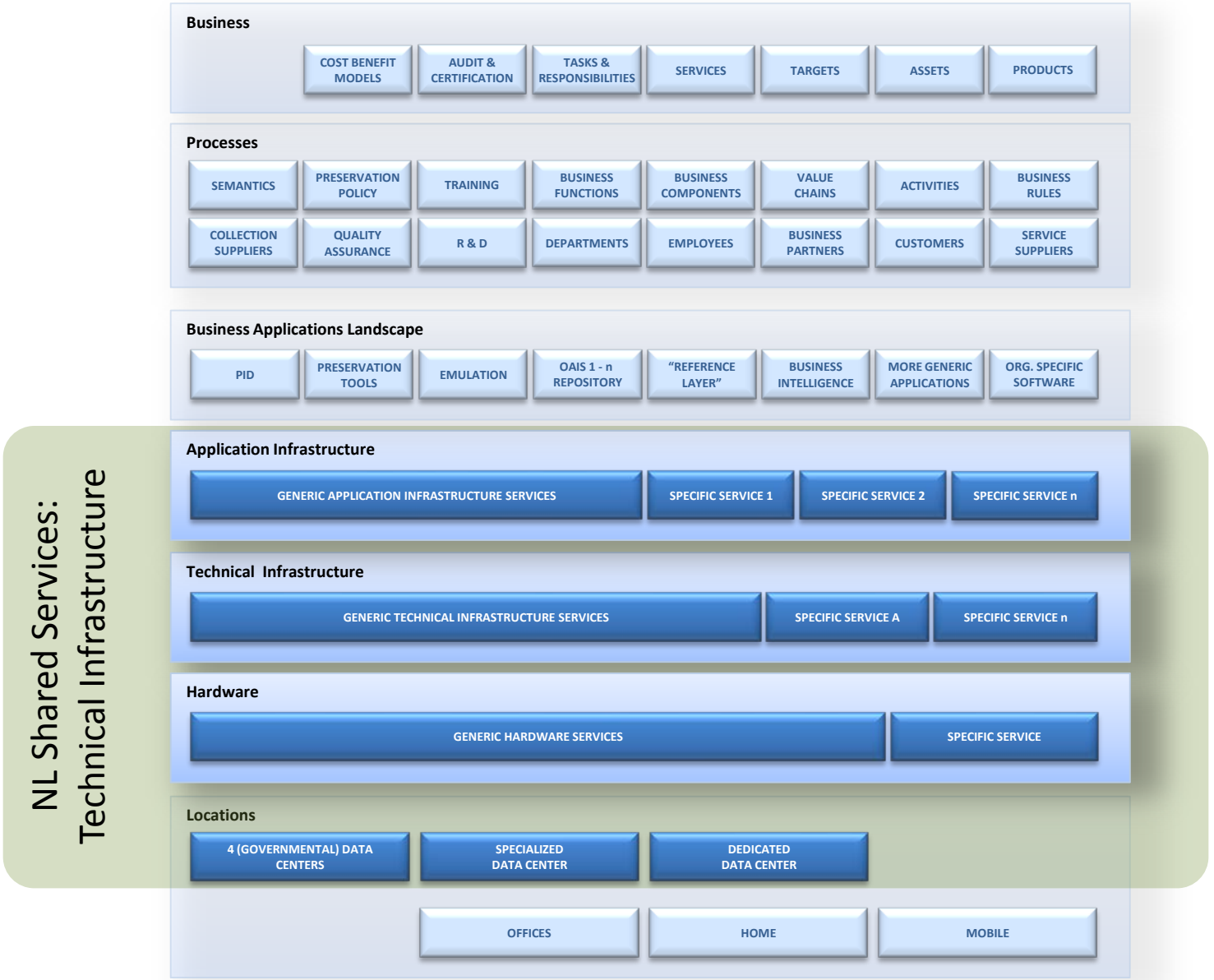
Scenario 3, DPI building blocks
(analytical framework taken into account)

NCDD
NATIONALE COALITIE DIGITALE DUURZAAMHEID

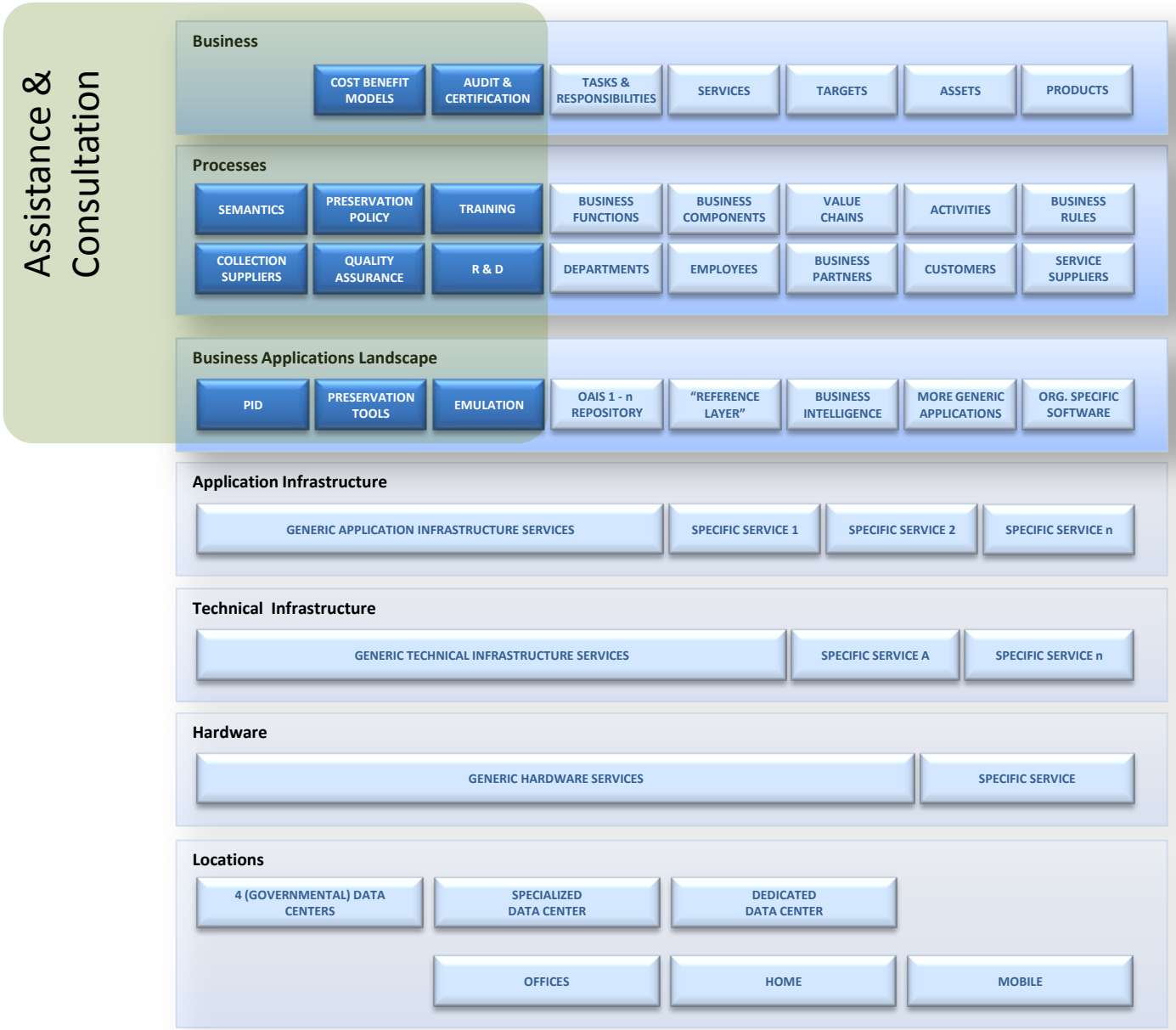


Scenario 3A, building blocks

NCDD
NATIONALE COALITIE DIGITALE DUURZAAMHEID

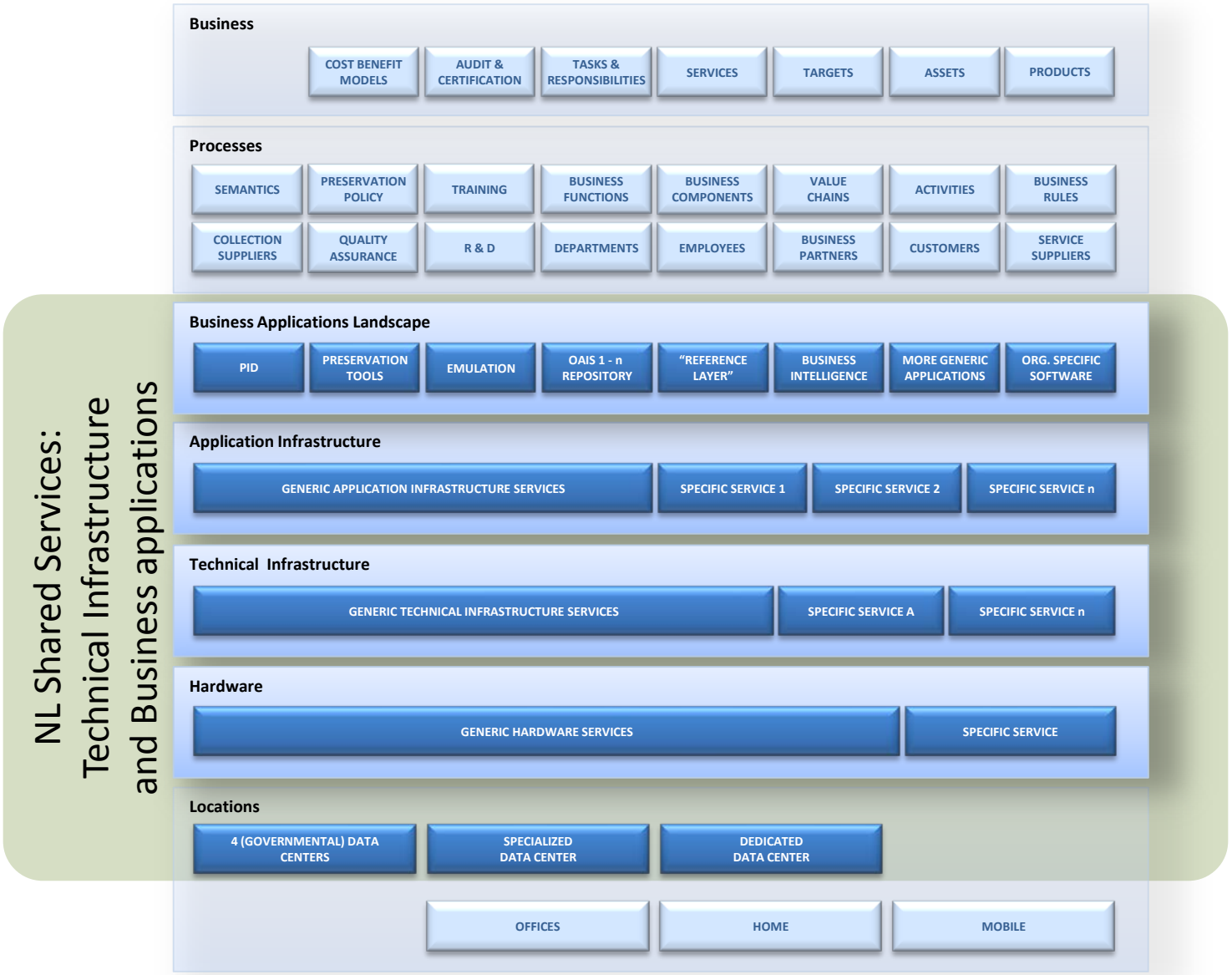


Scenario 3B, building blocks



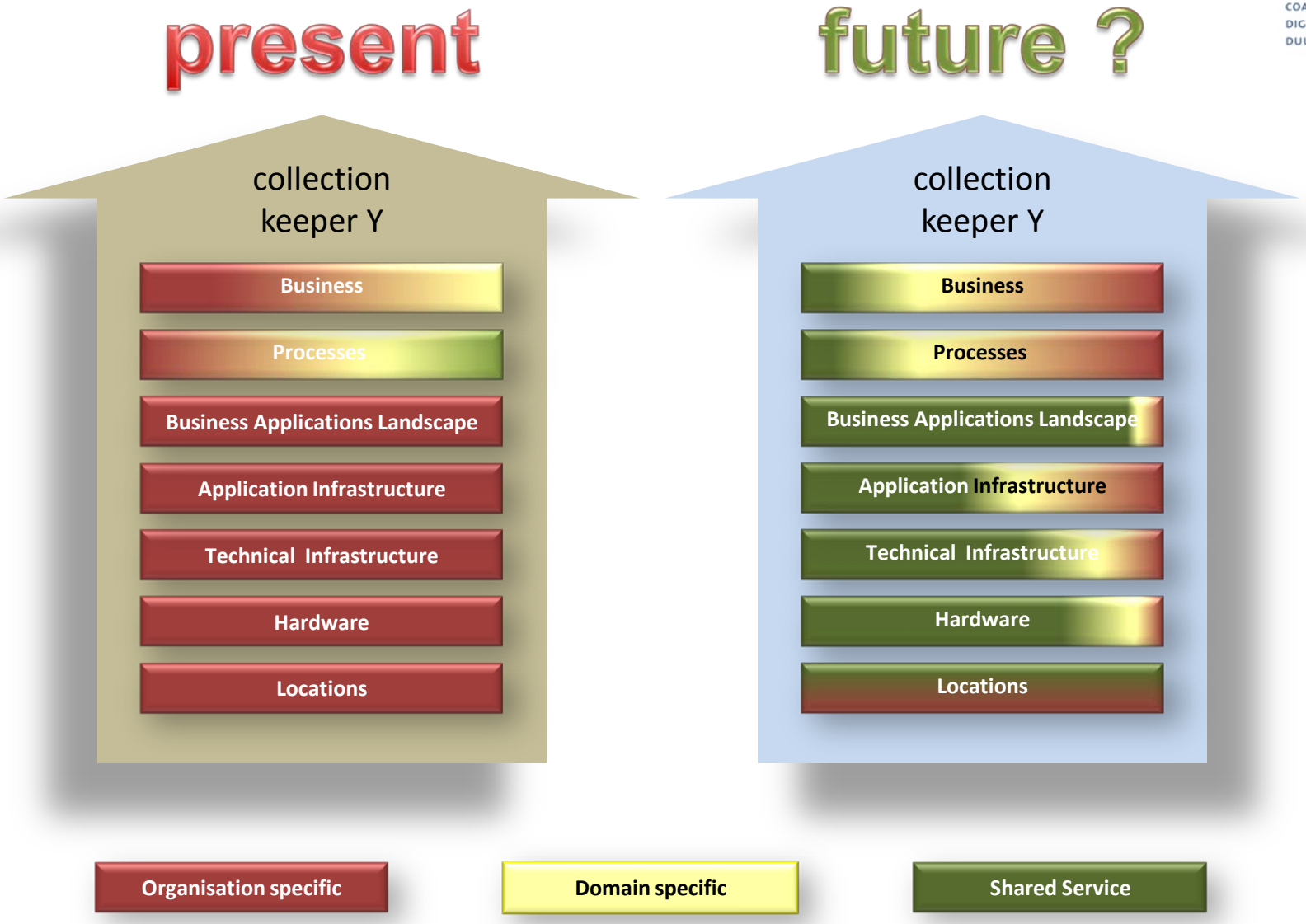
Scenario 3C, building blocks
(distributed national infrastucture = 3A + 3C)

NCDD
NATIONALE COALITIE DIGITALE DUURZAAMHEID



Scenario 3, present vs future

NCDD
NATIONALE COALITIE DIGITALE DUURZAAMHEID

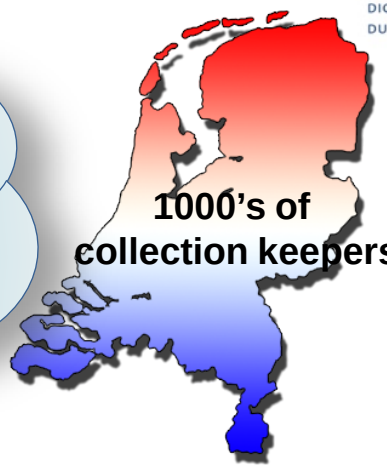


Scenario 3, future

NCDD
NATIONALE COALITIE DIGITALE DUURZAAMHEID

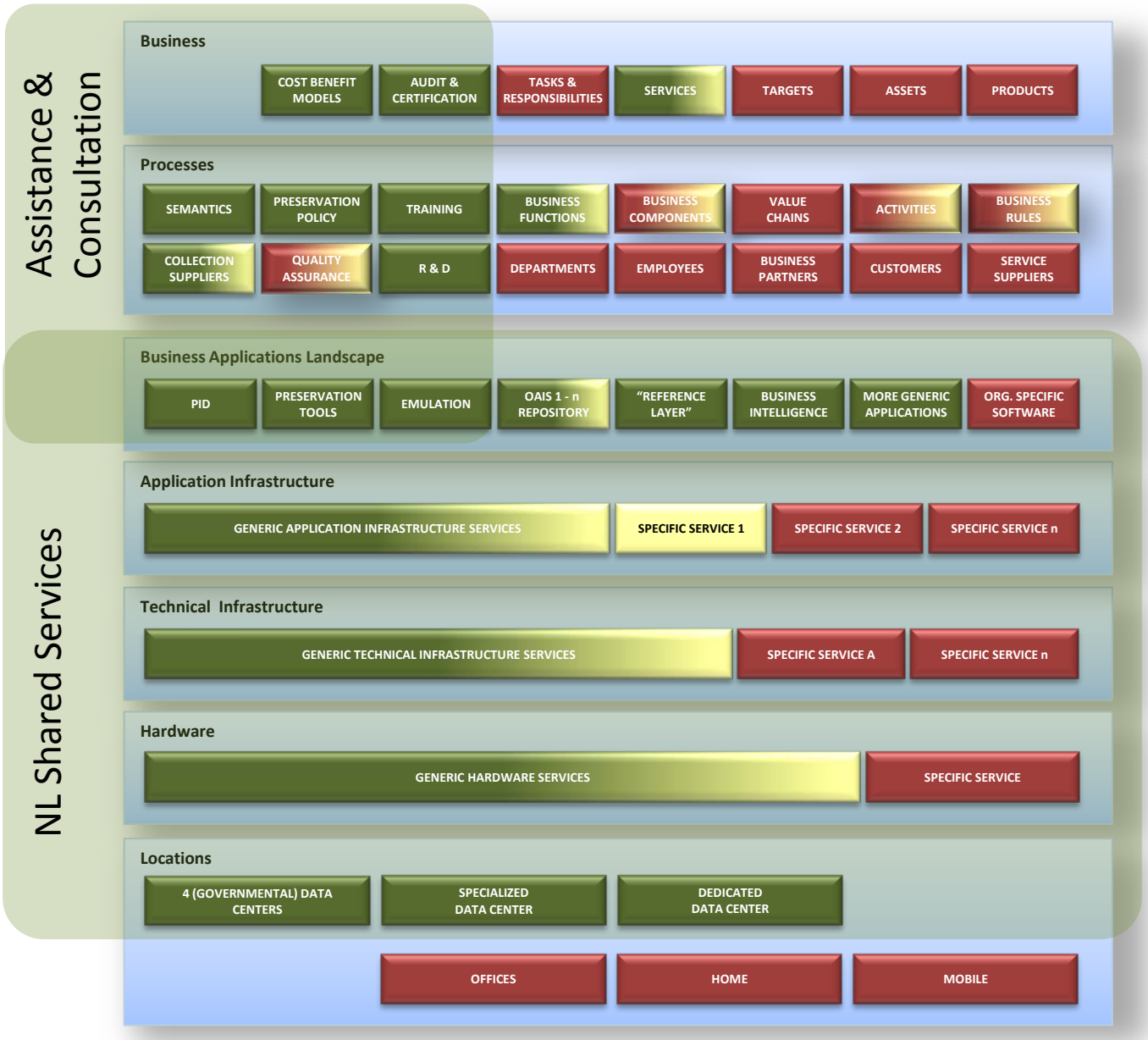
5 – 6
data centers

N : M



- Business
- Processes
- Business Applications Landscape
- Application Infrastructure
- Technical Infrastructure
- Hardware
- Locations

Scenario 3, in detail



NCDD
NATIONALE COALITIE DIGITALE DUURZAAMHEID

Organisation specific

Domain specific

Shared Service

Required next steps for NL-NDPI “2050”

NCDD
NATIONALE COALITIE DIGITALE DUURZAAMHEID

- Information Management Policy “Culture” by the Ministry of OCW input from NCDD seems required
- Based on this, develop the Mission & Vision for a Dutch National Digital Preservation Infrastructure (in line with the National Strategy)
- Develop a Collaborative Strategy
- Formulate long term Aims (20 – 30 years)
- Make the extended Business Case: e.g. elaborate on scenarios, expand on details, analysis of benefits, maturity analysis, cost – benefit calculations
- Formulate Governance
- Investigate, in a standardised way, the Requirements of all types of institutions and organisations involved (“Greatest Common Divisor” vs “Least Common Multiple”; it is the latter we need…)
- Define Goals (5 – 10 years)
- Define Objectives & Action Plan (horizon 1-2 years)
- Develop a Programme, start projects in line with National Strategy

Workshop

NCDD
NATIONALE
COALITIE
DIGITALE
DUURZAAMHEID

Agenda

NCDD
NATIONALE COALITIE DIGITALE DUURZAAMHEID

1. Introduction
2. NCDD

RESEARCH

3. Goals
4. Analytical framework
5. Matrix
6. Scenarios

WORKSHOP

7. Briefing
8. Work !
9. Plenary feedback

Workshop, briefing

NCDD
NATIONALE COALITIE DIGITALE DUURZAAMHEID

- Topics for workgroup 1 - 4:
  1. Is scenario 3 a sound one ?
     - missing elements ?
     - elements in the right layer ?
     - alternatives ?
  2. What are the pro's & con's of scenario 3 ?
  3. What are the risks, running scenario 3 ?
  4. Transition from "AS IS" to "TO BE" ?
     - time path ?
     - carrots & sticks ?
     - what is needed to achieve this ?
     - constraints (what must be in place to make it happen) ?

- Divide participants in 4 workgroups
- Please report your findings (flap over) in 45 minutes from now.

More information

NCDD
NATIONALE
COALITIE
DIGITALE
DUURZAAMHEID

www.ncdd.nl

http://www.ncdd.nl/blog/

http://groups.google.com/group/ncdd-nieuws

Program Manager: Marcel Ras

marcel.ras@ncdd.nl

+3170 31 40 180
+316 1477 76 71

Researcher: Joost van der Nat

jvdnat@ncdd.nl

+316 5204 8788