

The Goportis Digital Preservation Pilot Project Experiences made, lessons learned

Michelle Lindlar and Thomas Bähr
Future Perfect 2012

March 26th 2012, Wellington – New Zealand



The Goportis Pilot project

Conducted from October 2009 – October 2011

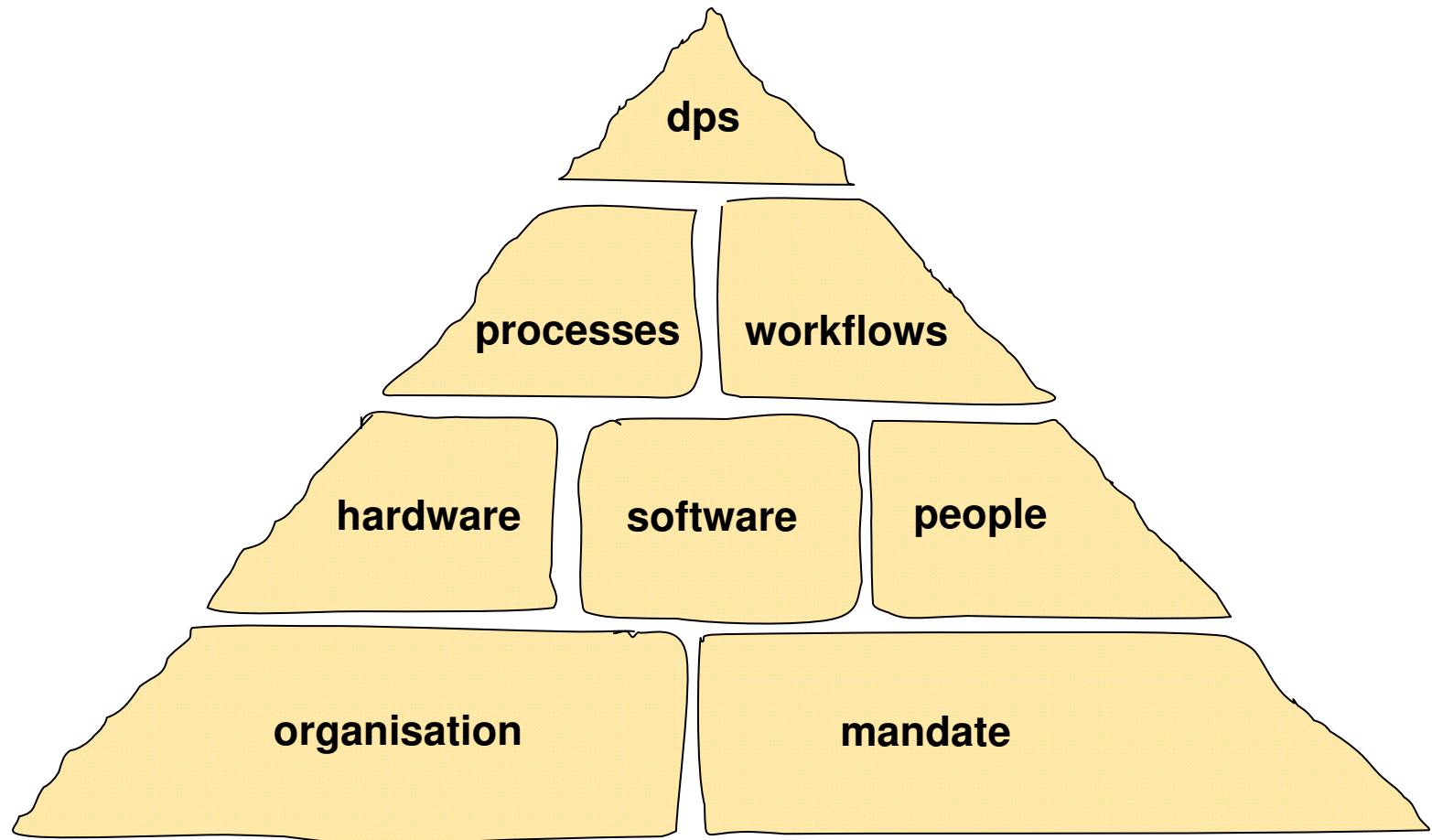
Goportis consists of the three German national subject libraries:
the German National Library of Science and Technology (TIB)
the German National Library of Medicine (ZB MED)
the German National Library of Economics (ZBW)

Goal: To determine and evaluate **technological, institutional and organisational** needs for a cooperatively operated digital preservation system.

Cooperatively operated means that all partners can work equally in the system.



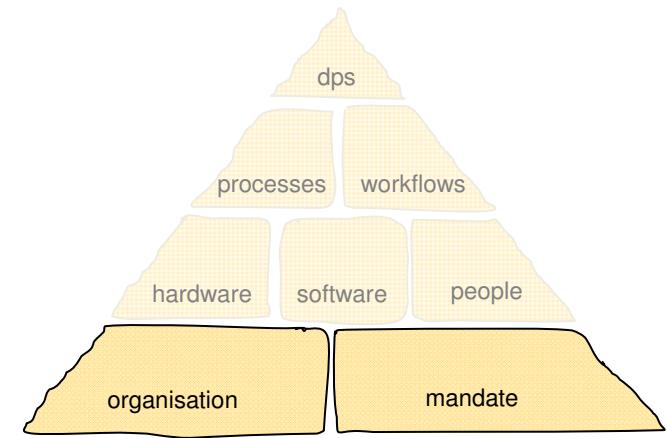
„Building a digital preservation system“



„Pyramid foundation“

Organisation

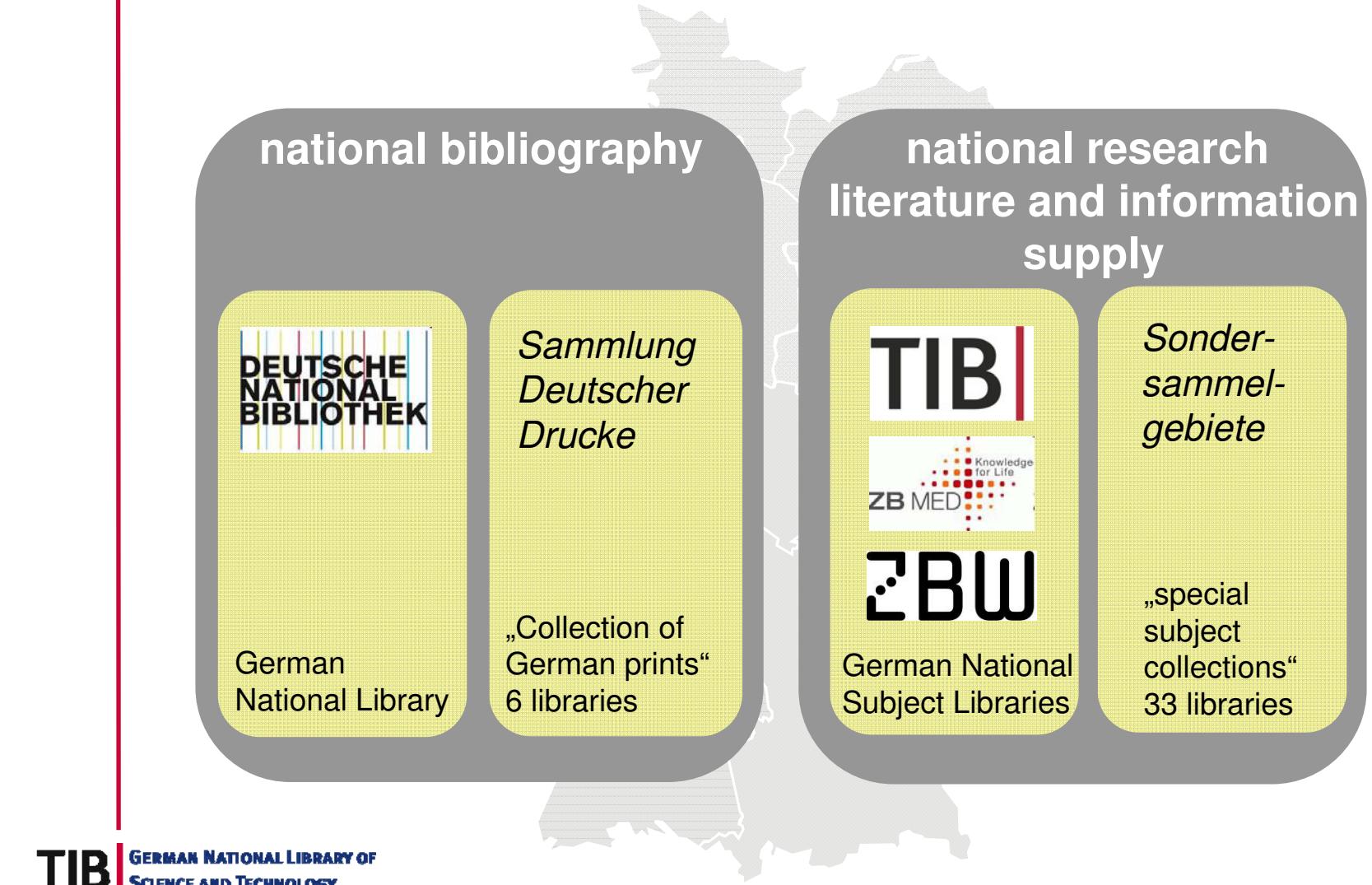
- type: library, archive, research institute, ...
- level: national, state, university, ...
- size: holdings, staff, budget, users, ...
- **defines your (national) position !**



Mandate

- Given by: act/law, superordinate organization/institution, self-given, ...
- For content: (sub)collection level, type of content, ...
- Including action: collecting, archiving, making available, ...
- **defines your (national) role !**

A distributed national (research) library system



Goportis



Subjects: engineering,
architecture, information
technology, chemistry,
mathematics, physics

Staff: 212

Holdings: 8.9 mio units



Subjects: medicine,
nutrition, environmental
science, agricultural
science

Staff: 122

Holdings: 1.6 mio units



Subject: economics

Staff: 239

Holdings: 4.4 mio units

different technical infrastructures (e.g. repositories, cataloguing systems)
different digital collections (e.g. AV, 3D objects)

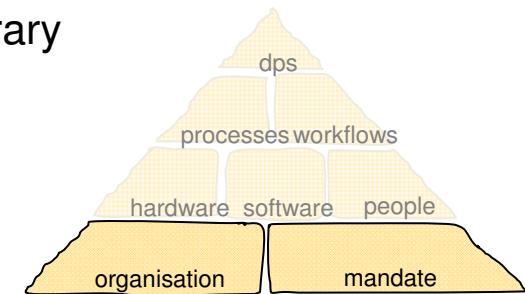
Pyramid foundation – lessons learned

One for all or all for one?

- Partner oriented model: almost identical organizations / mandates
- Service oriented model different forms of organizations / mandates
- More partners mean more complexity (communication, documentation, methods of operation)

Think about hierarchy within your institution

- digital preservation is a cross-functional task and an organisational change process
- during implementation phase it is beneficial to position the digital preservation in the hierarchy as close to library management as possible
- A permanent position of digital preservation within an institution will have to be found post-pilot



Pyramid 1st floor

Hardware / Infrastructure

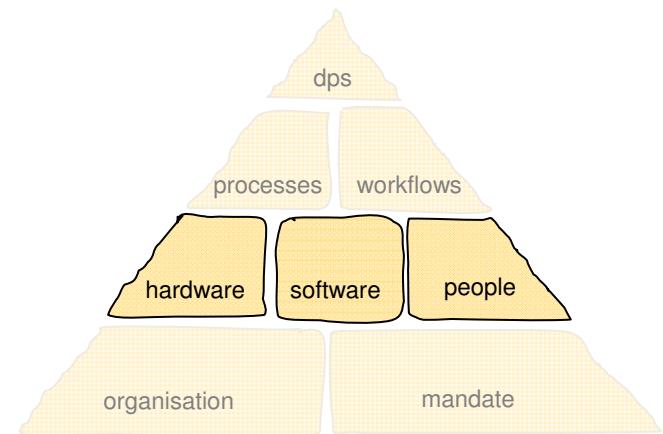
- Central or decentralized
- Open infrastructure ?
- Scalability and reliability

Software

- System or service
- custom-built or off-the peg
- commercial or open source

People

- size and structure
- Qualifications / knowhow
- Outsourcing possible?



System Choice: System vs. Service

System

- + control over your data
- + decision regarding actions
- + institutional/organizational needs can be met (flexibility)

- time from project start to roll out
- cost hardware / software
- more staff needed

Service

- + low staff costs
- + no hardware / software cost
- + time from project start to roll out

- no control over data
- actions based on service provider decisions
- access only in pre-defined cases

System choice: Off-the-peg vs. Custom build

Custom build

- + licensing cost low
- + modularity
- + quality
- + transparency

- + - community

- integration and development costs
- time from project start to roll-out
- support
- ongoing IT costs for development

Off-the-peg

- + lower IT resources development
- + continued development
- + time from project start to roll out
- + central end-to-end system

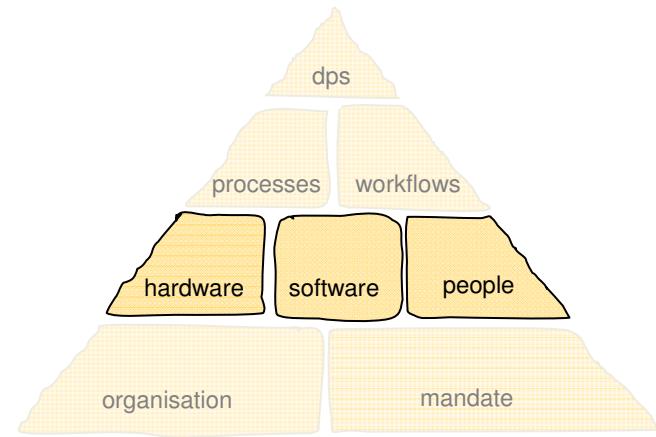
- + - support/service

- licensing cost
- integration of other systems
- dependency on company
- drawbacks in fulfillment of specific needs

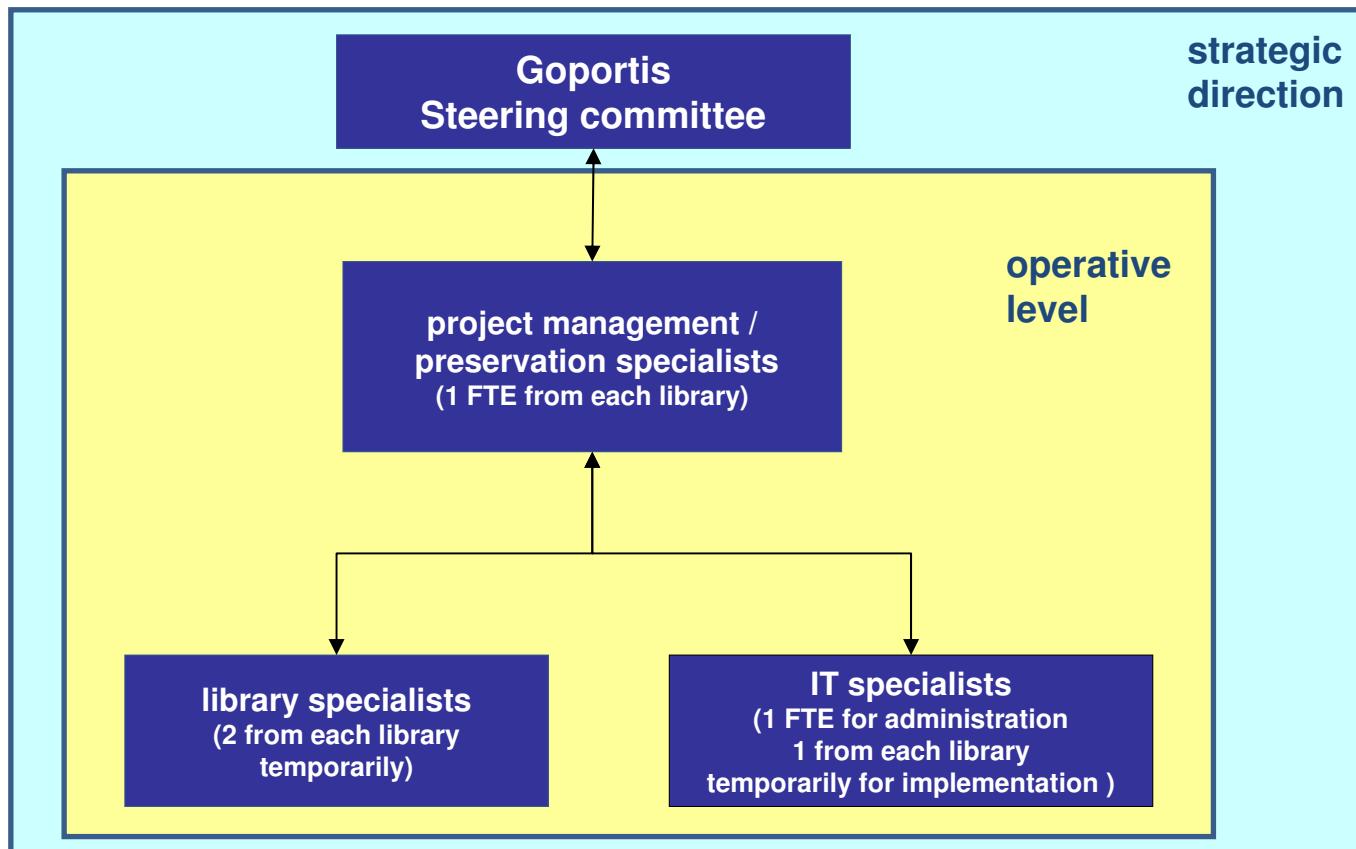
Pyramid 1st level – lessons learned „software“

Is the software ready for you?

- high value of a user community
- system is close to user needs –
but are the user needs YOUR needs?
- high value of a flexible system
(in regards to configuration, integration points, ...)
- clear exit scenario has to be defined



People – qualification / know-how



People – know-how

Preservation specialists

- Excellent understanding of formats, preservation procedures, risks, ...
- Good understanding of workflow procedures
- Basic understanding of IT procedures

Library staff

- Good understanding of digital preservation
- Experts for one or more workflows
- Experts for descriptive metadata (DC, MARC, MAB, ...)

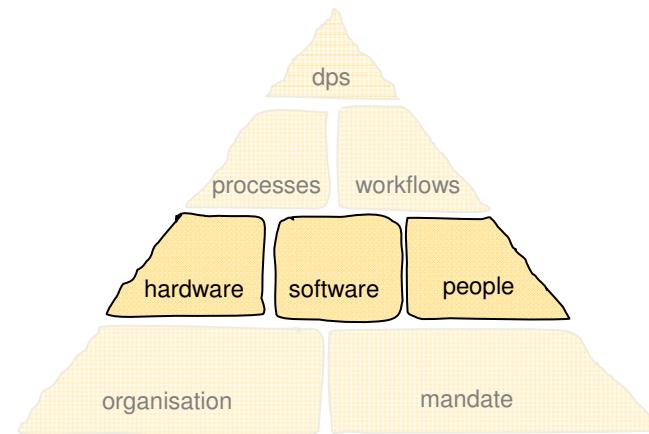
IT specialists

- Good understanding of digital preservation
- Programming skills
- Database expert

Pyramid 1st level – lessons learned

Are you ready for digital preservation?

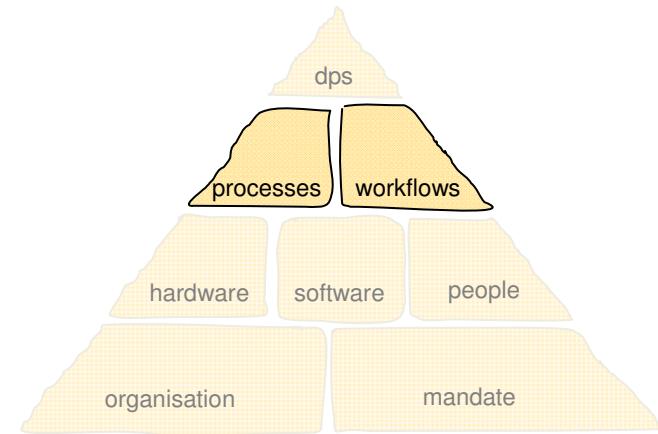
- „Know-How“ is a continuous process
- three pillars of knowledge: library processes, IT, preservation practise
- “spread the word“ within your institution !
- “community sourcing“



Pyramid – 2nd level

Processes

- Specific tasks within your institution related to preservation
- Organizational process
- Technological process
- Can involve humans and/or systems
- **Community building**



Workflows

- Combination of tasks/processes to form a meaningful chain
- In library context „workflows“ usually pertain to handling materials (or users)
- Can involve humans and/or systems
- **Traditional workflows**
- **Digital workflows**

Processes – Community Building

Value of Communities for digital preservation

- we all have similar problems
- „universal“ knowledge regarding formats, risks
- new developments often part of „projects“
- „keeping tools alive“
- standardization for digital preservation

Contributions of Communities – a few examples

- DPC Technology Watch Reports <http://www.dpconline.org>
- OPF Blogs and Wiki <http://openplanetsfoundation.org/>
- KEEP public deliverables <http://www.keep-project.eu>
- DPOE workshops <http://www.digitalpreservation.gov/education/>
- nestor working groups
<http://www.langzeitarchivierung.de/eng/index.htm>
- PREMIS standard <http://www.loc.gov/standards/premis/>

Processes – Goportis Community Involvement



„International level“



„National level“



„Institutional level“



„Procedural level“

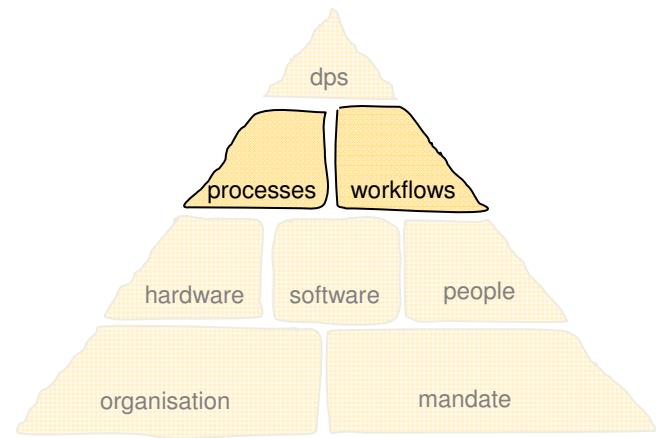


„product/tool level“

Pyramid 2nd level – Lessons learned „processes“

Processes – Community Building

- community involvement is the process to keep your know-how up to date
- think about the level of community involvement right for you (related to organization structure)
- try to plan how much time you can spend on community activities
- never underestimate the role of institutional level communities !



Workflows – Traditional vs. Digital workflows

Traditional workflows (e.g. cataloguing, selection for acquisition)

- basis of library procedures
- handling materials throughout their lifecycle in the library's holdings
- often static
- always require human interaction

Digital workflows (e.g. ingest, risk management)

- configuration within a digital system
- handling digital objects throughout their lifecycle in the library's digital system(s)
- changes in the system may require changes in the workflow
- may be automated

Workflows – Ingest in the Goportis Pilot Project

manual ingest („dissertations“)

Files are loaded into Rosetta by librarian



Librarian enters minimal set of descriptive Metadata



Objects are „validated“ (identified, characterized, virus check, etc.)



Problems in the validation process need to be solved by preservation specialist



Objects are manually linked with cataloguing system



Objects are double-checked, „approved“ and passed to archival storage

automated ingest („repository“)

Files are picked up by Rosetta from a predefined directory



Minimal set of descriptive metadata is supplied by repository with file



Objects are „validated“ (identified, characterized, virus check, etc.)



Problems in the validation process need to be solved by preservation specialist



Objects are automatically linked with cataloguing system

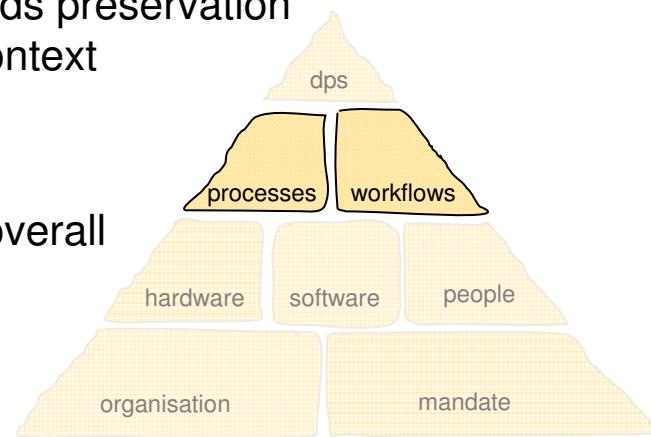


Objects are passed to archival storage

Pyramid 2nd level – Lessons learned „Workflows“

Workflows – Traditional vs. Digital

- is the main difference between „traditional“ and „digital“ a move towards automation?
- automation is not always a technical problem
- good understanding of benefits and drawbacks of automated processes/workflows
- think about your institutional approach towards preservation and what should not be automated in that context
- just because something can be automated, should it be?
- your workflows need to be in-line with your overall archival policy
- define which sources you trust and why !



You think you're done? Forget it!

Organization

position digital preservation as a fixed unit / department / ... in your institution

Mandate

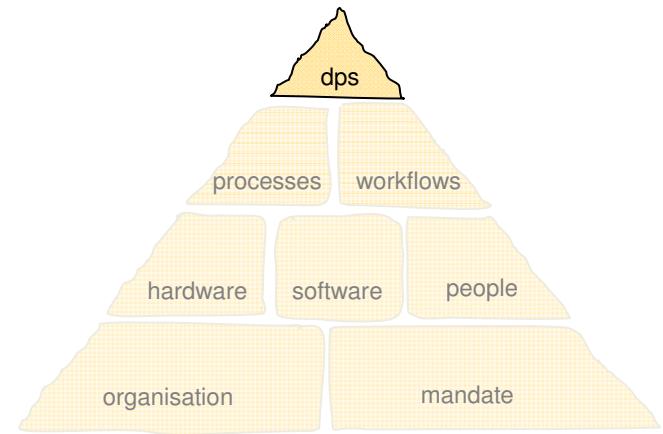
compare your mandate to your digital preservation strategy and to the legal situation

Hardware / Infrastructure

plan ahead for scalability and consistently check your reliability procedures

Software

consistently check your exit strategy; look for tools to help you with different preservation tasks (e.g. migration tools)



You think you're done? Forget it!

People

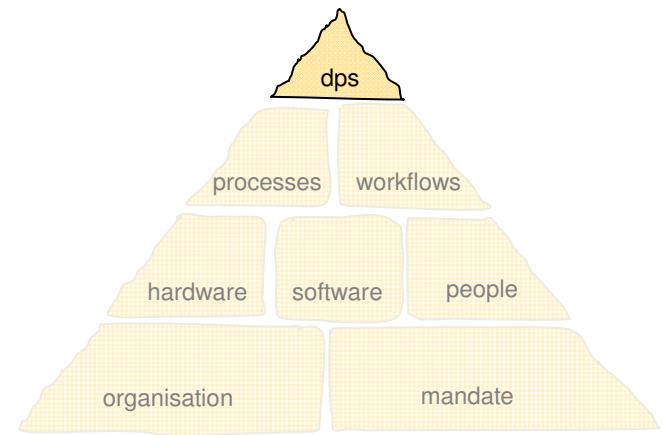
Include digital preservation as a fixed part of the work description of all staff involved – on paper and in their heads!

Processes

Integrate community activities as a fixed slot in your institution.

Workflows

Integrate more collections into your digital preservation system. Find the right balance between traditional and digital/automated workflows for your institution/your collections.



Collaboration is a key to success!



Thank you for your attention!

