

Creating a Framework for applying OAIS to Distributed digital preservation

Sam Meister

Educopia Institute

2º Seminário de Preservação Comum de Património Digital

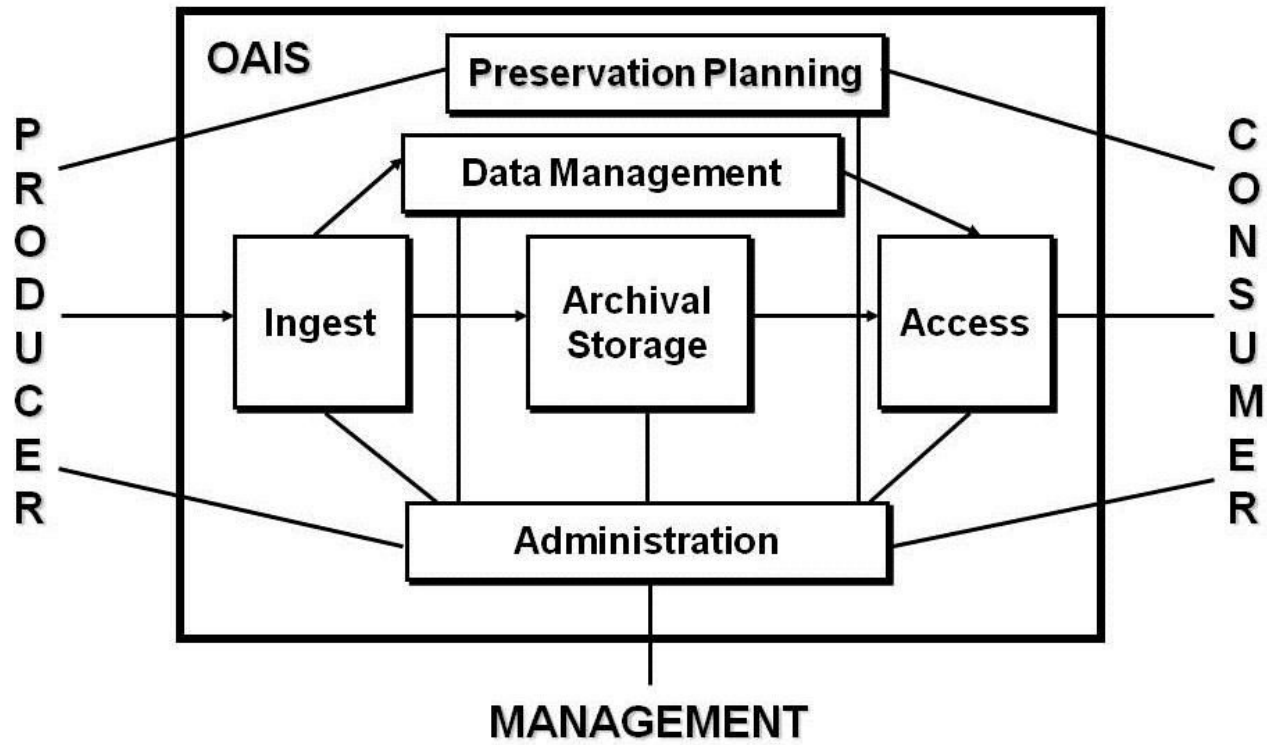
Lisbon, Portugal

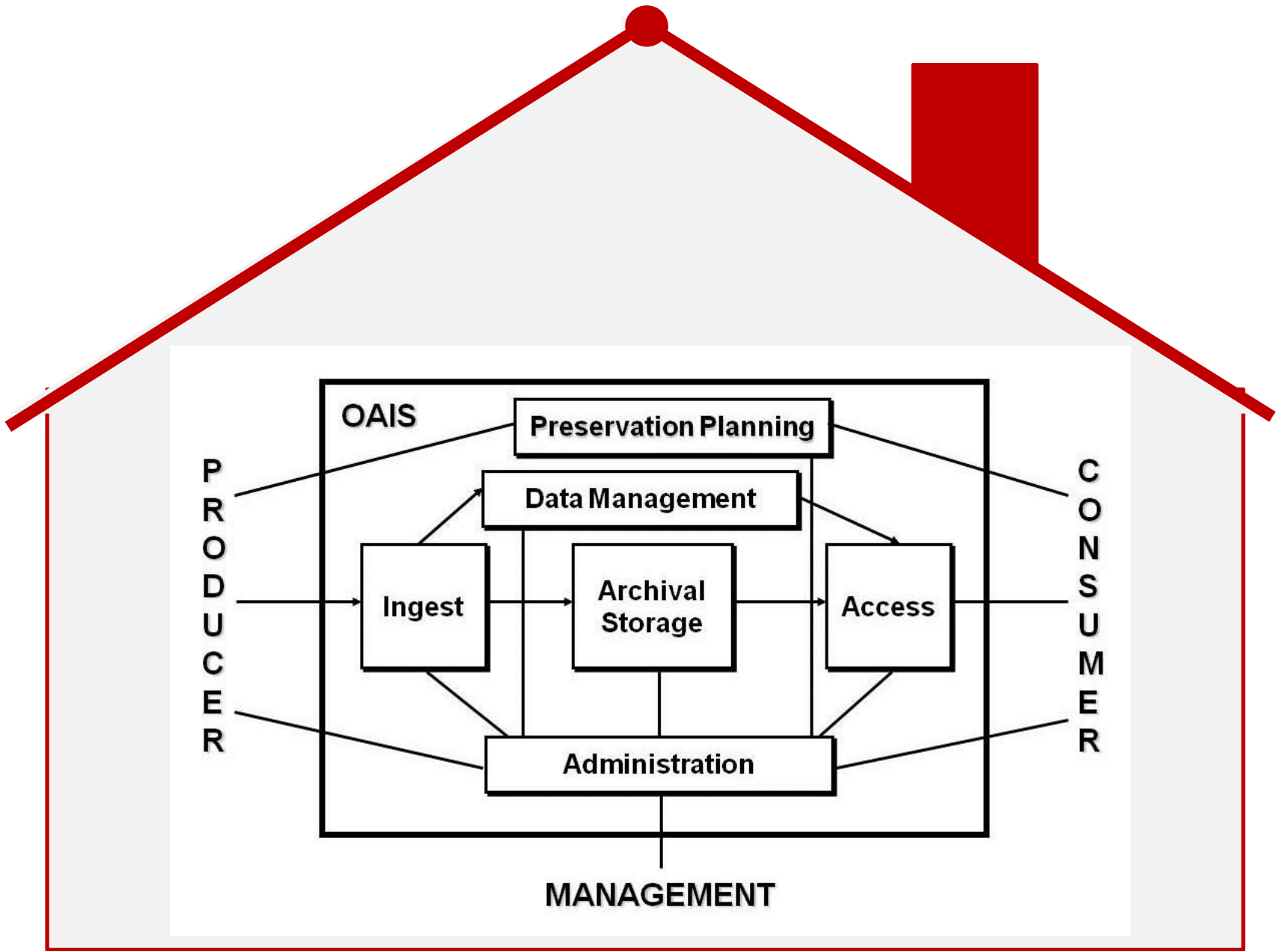
June 21-22, 2016

www.educopia.org

EDUCOPIA
INSTITUTE

Need / Motivation





OAIS

Preservation Planning

Data Management

Ingest

Archival Storage

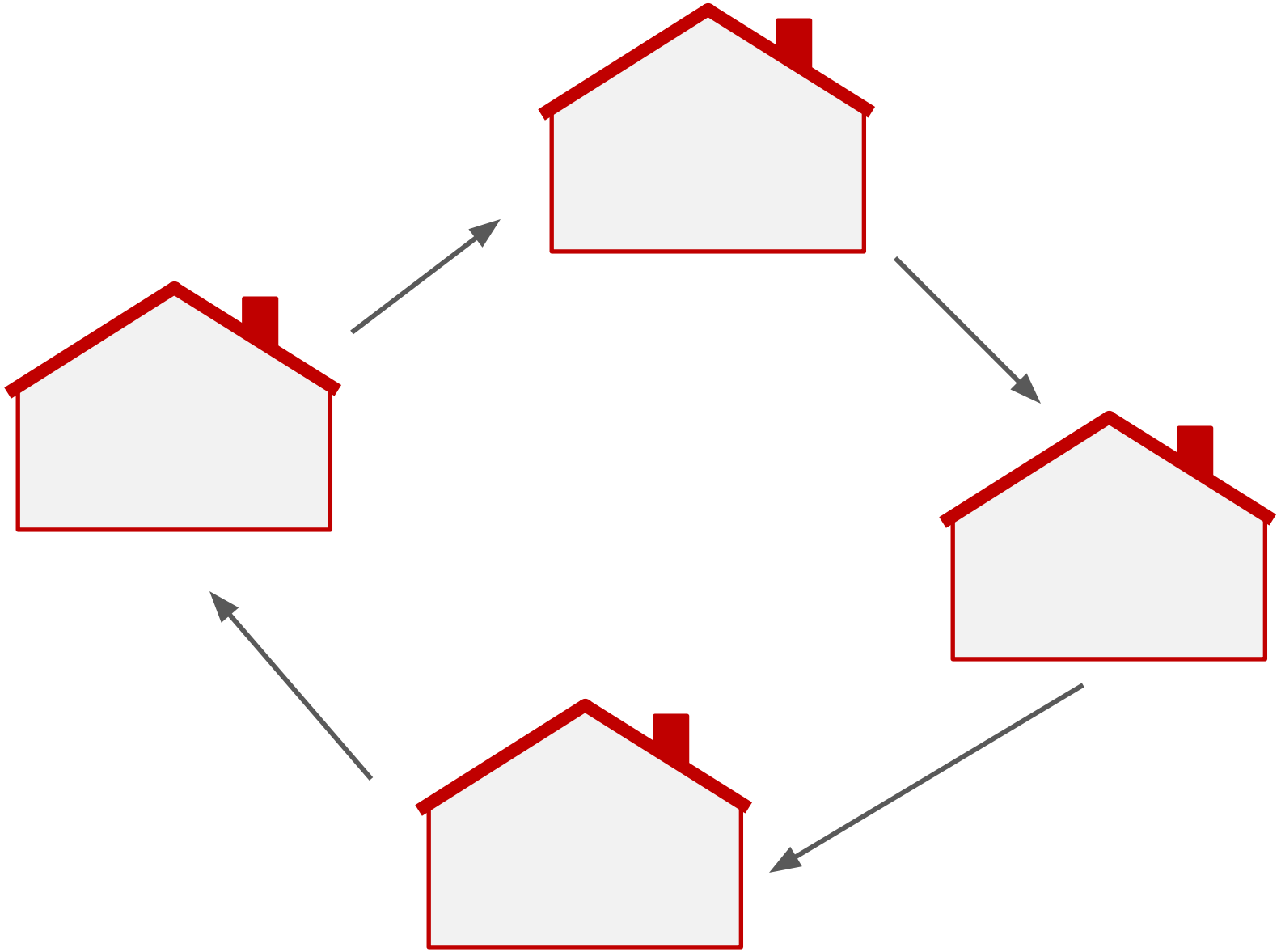
Access

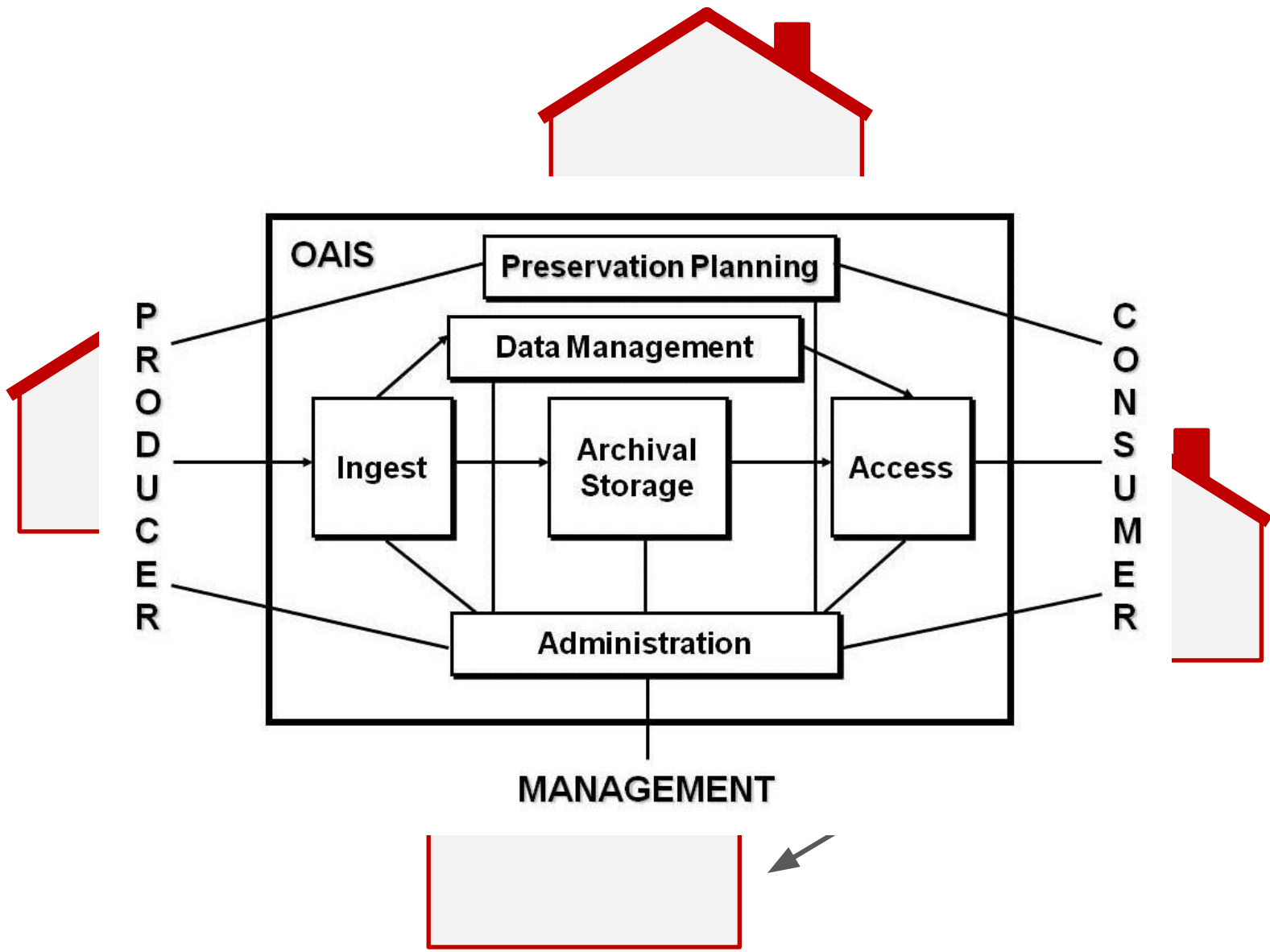
Administration

PRODUCER

CONSUMER

MANAGEMENT





Practitioners performing digital preservation
in highly distributed ways need methods
to **describe, analyze and audit**
Distributed Digital Preservation (DDP)

Existing DDP approaches lack common **vocabulary**

Need **models** for building effective, reliable, and auditable distributed preservation environments

Need understanding of DDP **roles and responsibilities**

Need means to **audit** distributed environments

Goals

Create a
Framework for Applying the Reference
Model for an Open Archival Information
System (OAIS)
to
Distributed Digital Preservation (DDP)

To increase **understanding and awareness** of the distributed digital preservation process amongst:

- Practitioners using preservation solutions
- Organizations seeking preservation solutions
- Providers of (parts) of preservation solutions
- Organizations needing audits of DDP solutions
- Organizations doing audits of DDP solutions

Audience

Organizations that are seeking to jointly develop or enhance DDP systems

Auditing bodies that are seeking to evaluate DDP systems

Producers / Consumers that are seeking to deposit and/or access their digital content in such systems and seeking to understand their operating principles

Project partners



Timeline

2011 - Formal Working Group convened to document the elaborations and methods needed to assist in applying OAIS to the DDP environment

2012 - White paper released

2013 - Presentations and community feedback

DDP Framework

Terminology

**Auditing
methodologies**

**DDP
Framework**

**Roles &
responsibilities**

Models

Terminology

Distributed digital preservation

“use of replication, independence, and coordination to address the known threats to digital content through time to ensure their accessibility”

Distributed digital preservation

*“use of **replication**, independence, and coordination to address the known threats to digital content through time to ensure their accessibility”*

Distributed digital preservation

*“use of **replication**, **independence**, and coordination to address the known threats to digital content through time to ensure their accessibility”*

Distributed digital preservation

*“use of **replication**, **independence**, and **coordination** to address the known threats to digital content through time to ensure their accessibility”*

Administrative unit:

a unit that can cover administrative tasks like common watch over the network, technical coordination, common planning and maintenance

Replica unit:

a unit that forms the basic storage for a copy of data in form of bit streams

Processing unit:

a unit that can do computer based processing on digital contents

Roles and responsibilities

Administrative

- Negotiate and review content
- Coordinate and manage software and hardware configurations at multiple institutions
- Establish common standards and policies

Preservation planning

- Monitor the archive's replicated holdings and technology environment to determine whether specific preservation actions are needed
- Review and update standards and policies

Ingest

- Establish independent yet replicable SIP data models
- Replicate SIPs to multiple, geographically distributed and separately-hosted archive nodes
- Ensure that descriptive and administrative data elements are properly collected, recorded and available to each partner or stakeholder

Archival storage

- Obtain AIPs from one source organization and ensure replications are synchronized
- Coordinate to achieve health and parity of storage availability to ensure adequate replication according to policy
- Perform content integrity audits at appropriate intervals across replicated copies and synchronize/repair copies between distributed holdings as deemed necessary

Data management

- Generate reports from distributed administrative data
- Replicate data management services effectively to ensure continuity of service
- Effectively manage and maintain the Archive database(s) for the purposes of discovery and producing usable reports regarding Archive holdings

Access

- Establish standard and secure protocols and permissions for identifying and retrieving DIPs from one or more independent institutions that are performing replication
- Provide for one or more shared and hosted dissemination location(s) where DIPs can be prepared and provided to end-users

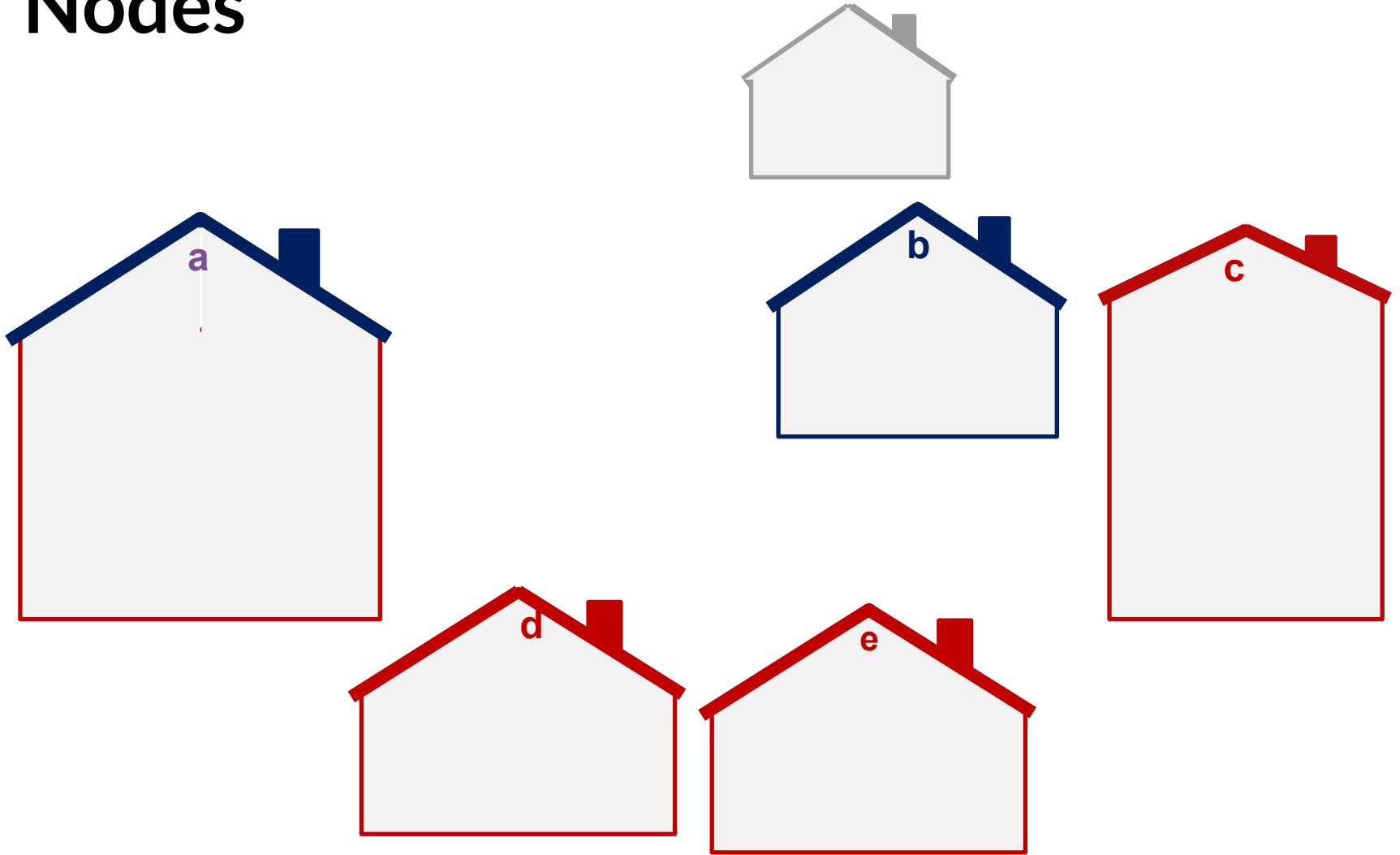
Models

Purpose of Models

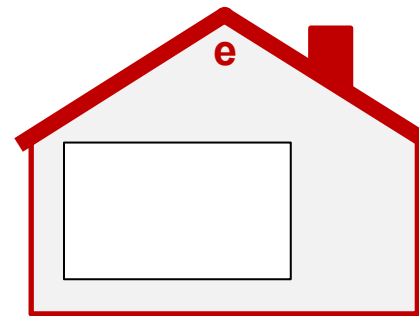
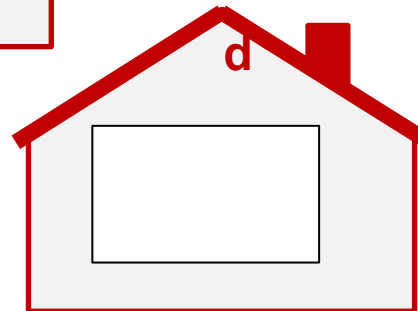
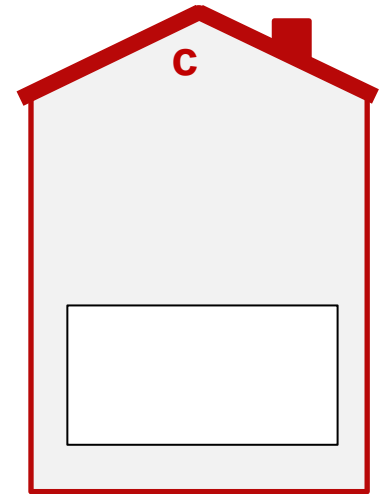
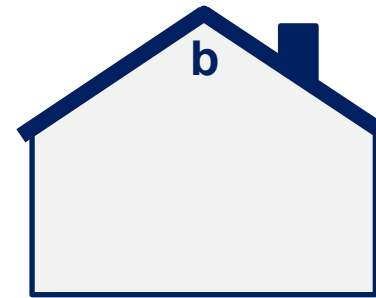
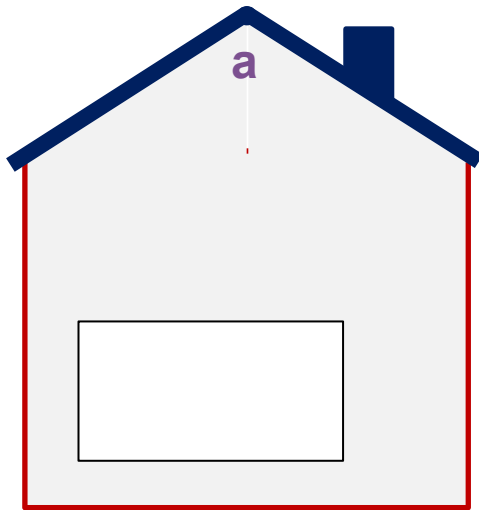
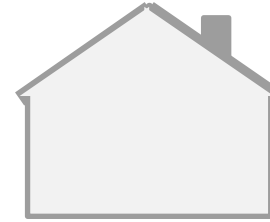
- To overview the elements that can be important for a specific case / scenario
- To understand and analyse the distribution in a specific case
- To select the nodes and units that can need special attention in analysis and audits

DDP Model

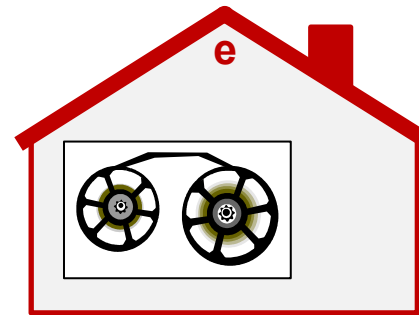
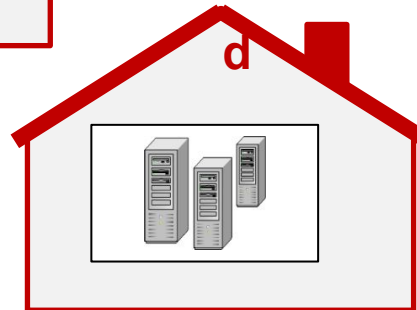
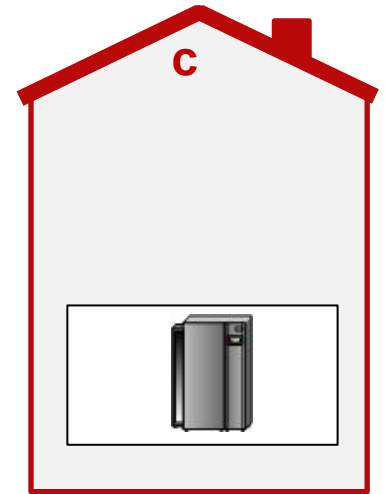
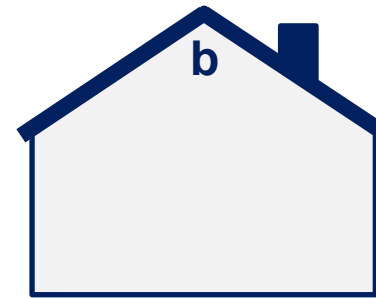
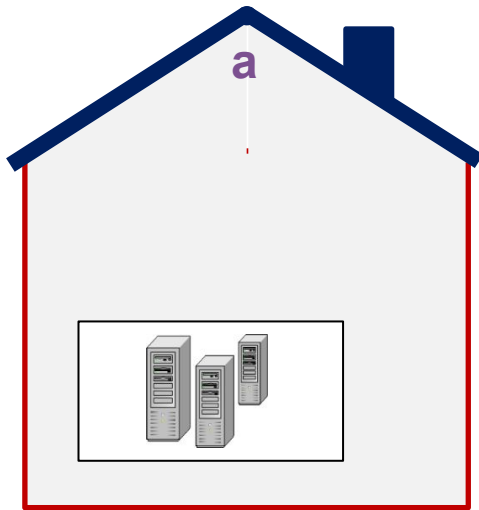
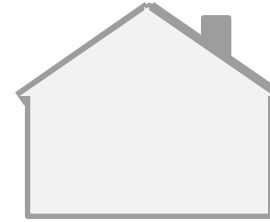
Nodes



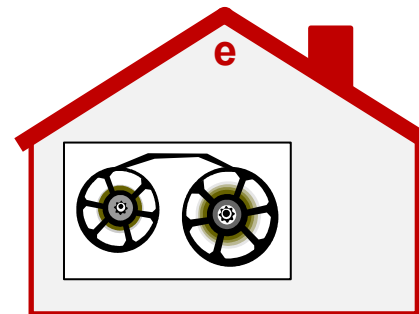
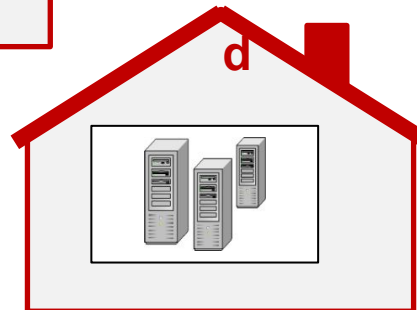
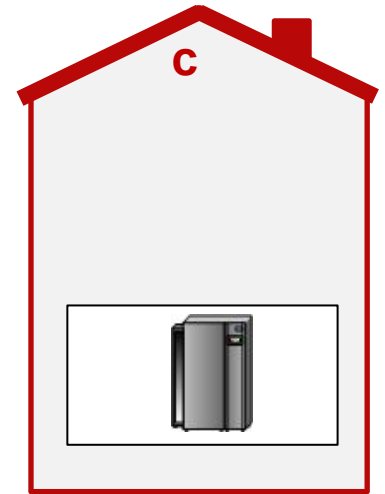
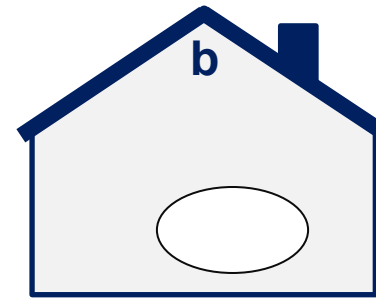
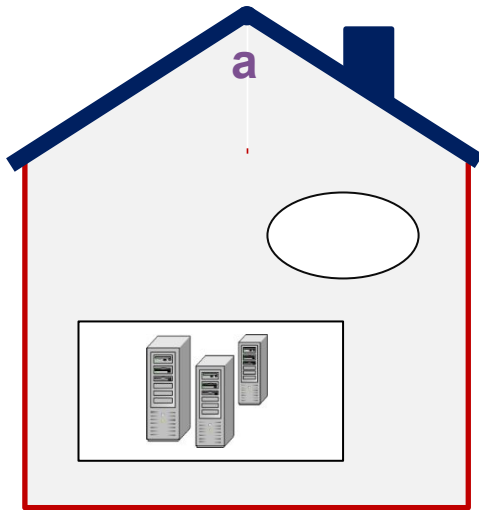
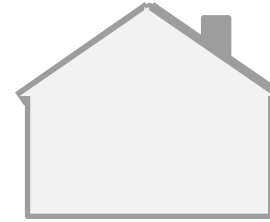
Replica units



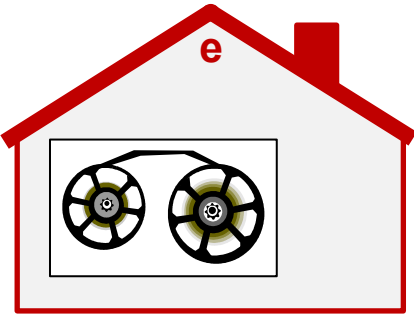
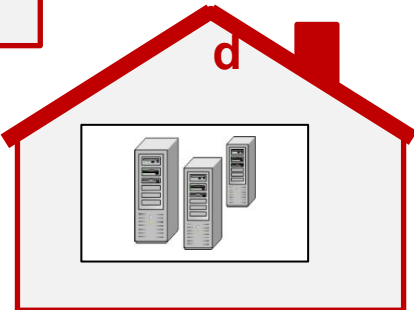
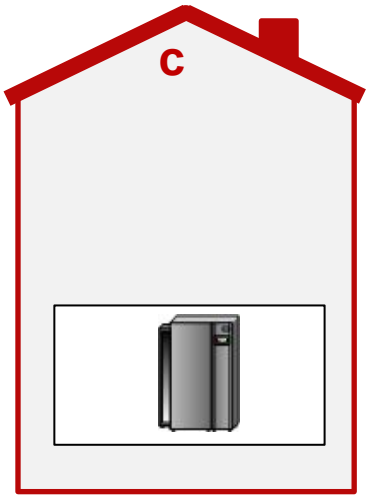
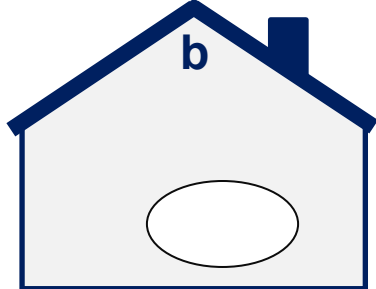
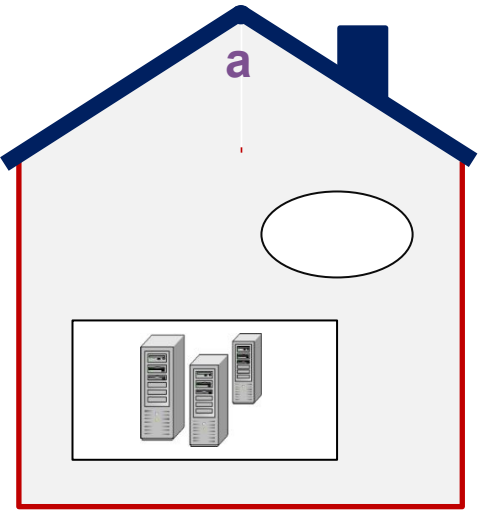
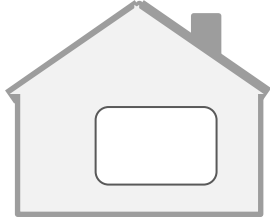
Replica units



Processing units

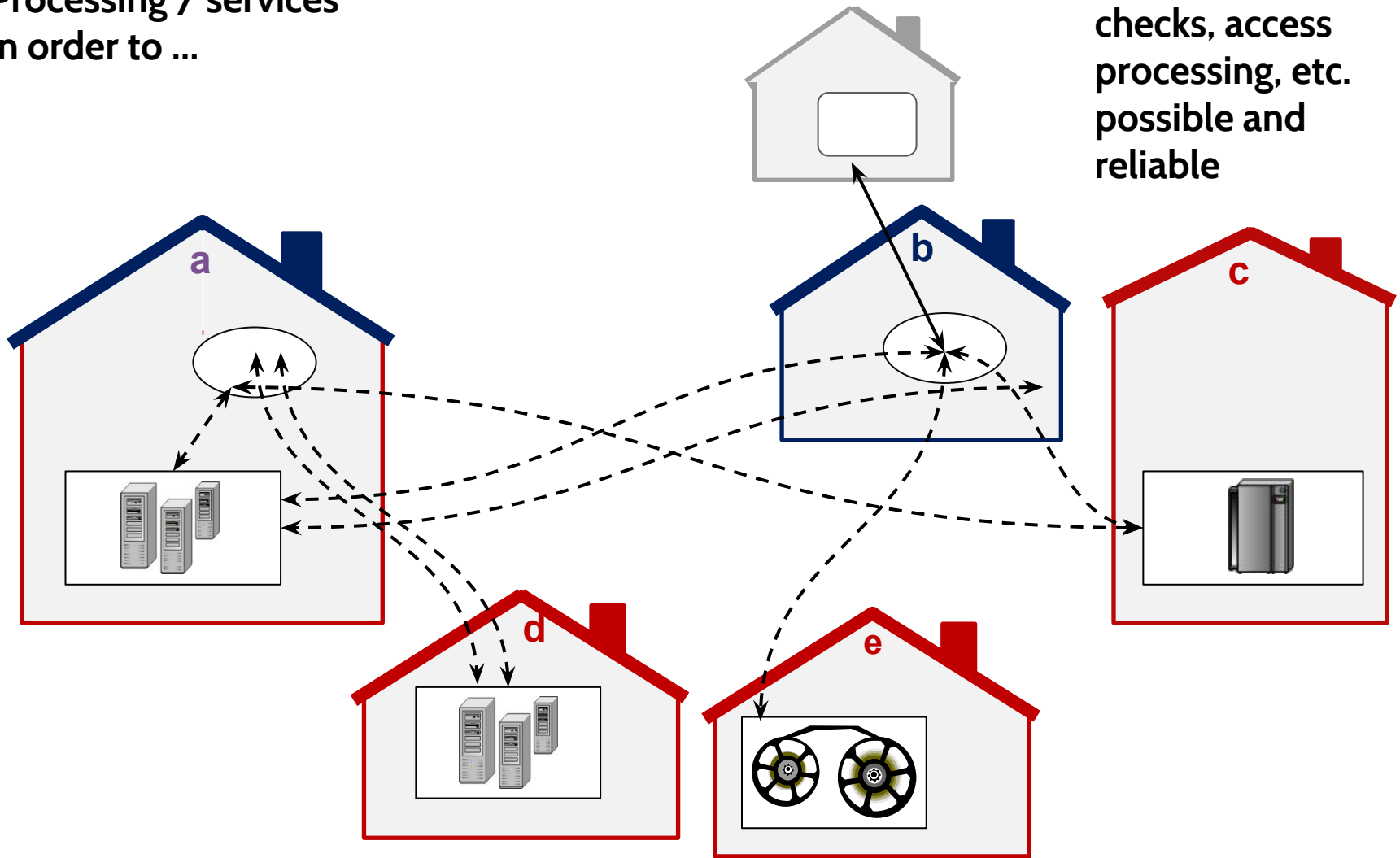


Admin units



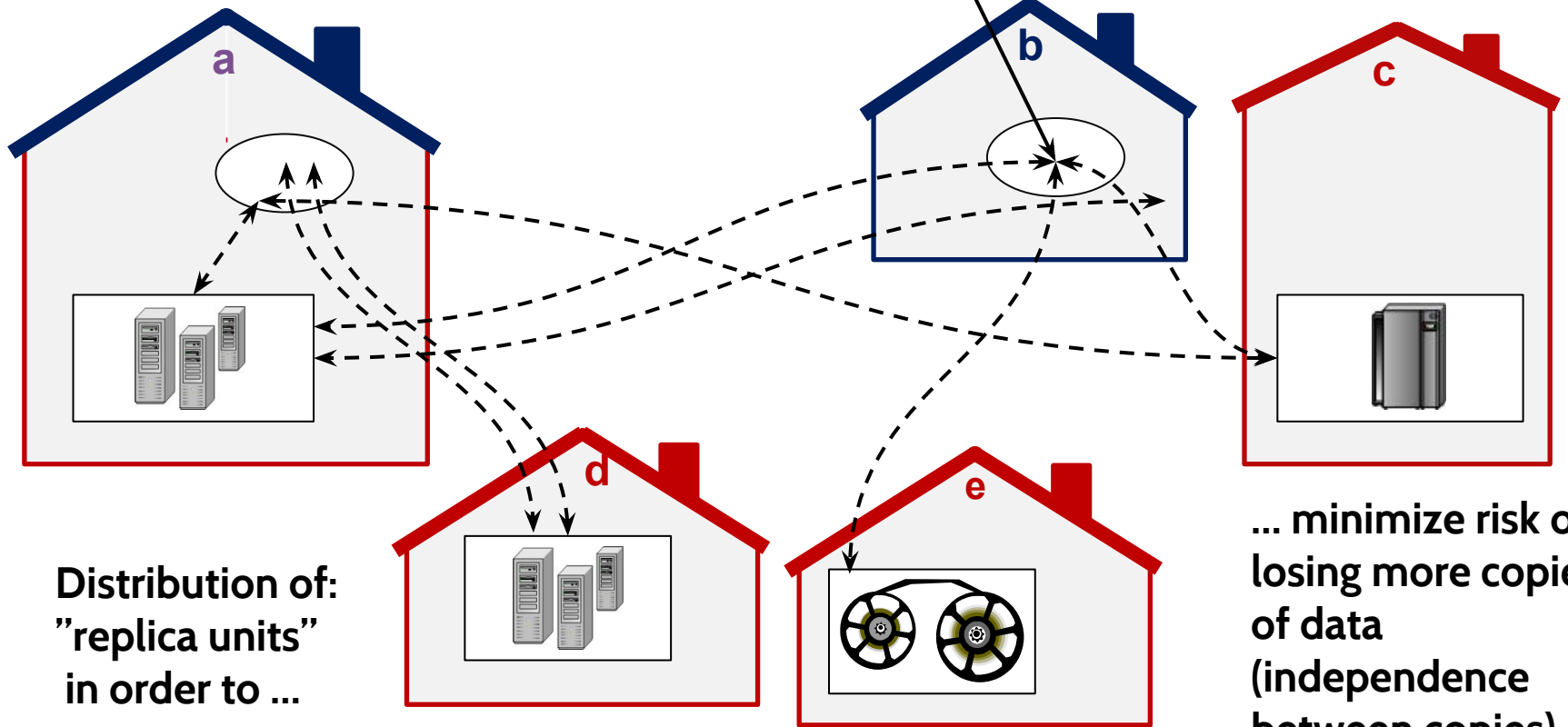
Distribution of:
Processing / services
in order to ...

... make Ingest,
checks, access
processing, etc.
possible and
reliable



Distribution of:
Processing / services
in order to ...

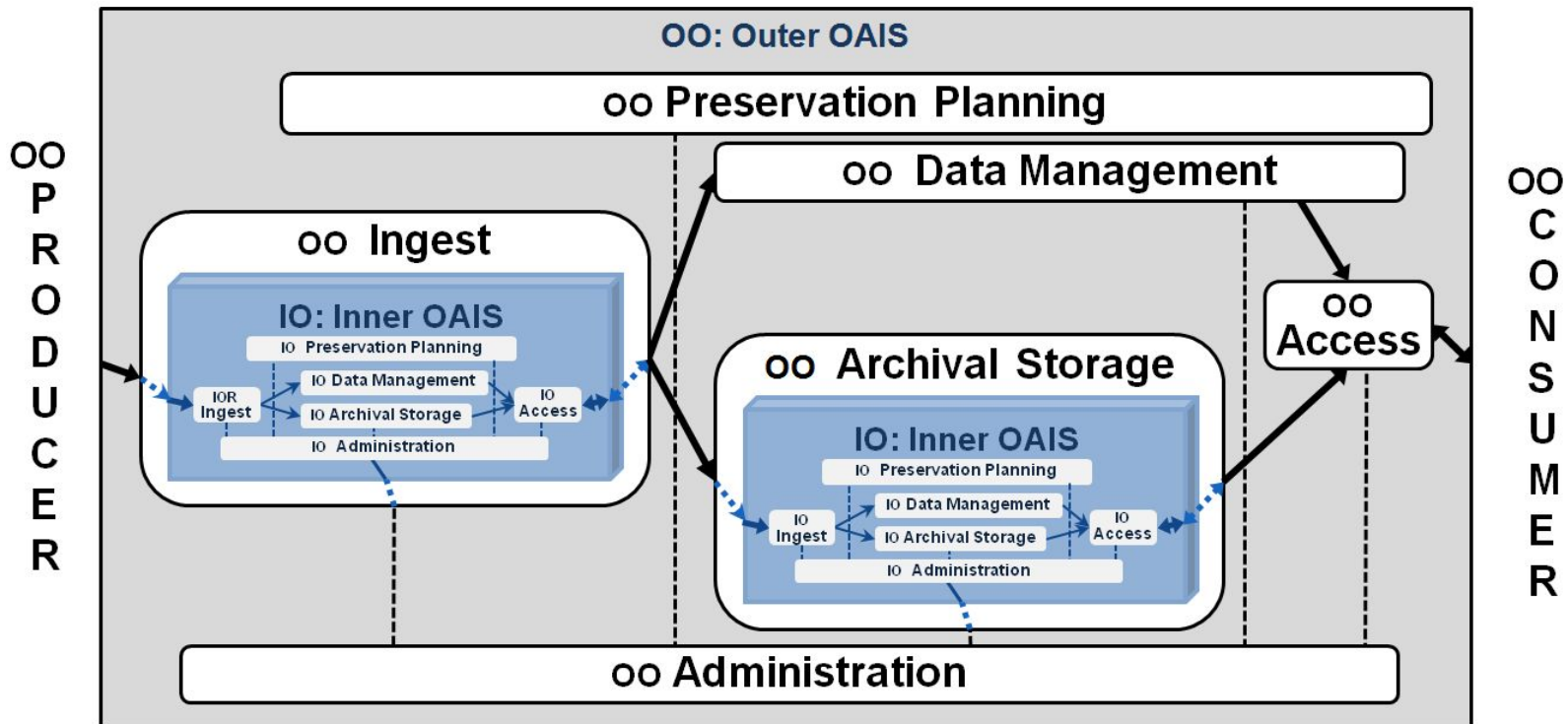
... make Ingest,
checks, access
processing, etc.
possible and
reliable



Distribution of:
"replica units"
in order to ...

... minimize risk of
losing more copies
of data
(independence
between copies)

Outer OAIS - Inner OAIS Model



Auditing Methodologies

Audit trends

Provide an overview of state of the art auditing methods and challenges

How to approach audits of DDP

Address possible ways to audit DDP environments

How to approach audits of DDP

Address how audit metrics from TDR/ISO 16363 and DRAMBORA can be interpreted for DDP environment

Future

Continued maturity of distributed digital
preservation networks

OAIS Review (2017)

Resources

Project website

<https://educopia.org/research/ddp>

Zeirau, E. & Schultz, M. (2013) Creating a framework for applying OAIS to distributed digital preservation

<https://educopia.org/presentations/creating-framework-applying-oais-distributed-digital-preservation>

Zierau, E., & McGovern, N. Y. (2014). Supporting the analysis and audit of collaborative oais using an outer oais-inner oais (oo-io) model. Preservation (DDP), 9(5)

http://digitalbevaring.dk/wp-content/uploads/2014/12/Zierau_McGovern_Outer_Inter_OAIS.pdf

DDP Exemplars

Archivemata (<https://www.archivemata.org/>)

Chronopolis (<https://libraries.ucsd.edu/chronopolis/>)

COPPUL (<http://coppullockssgroup.pbworks.com>)

Data-PASS (<http://www.data-pass.org/>)

DuraCloud (<http://www.duracloud.org/>)

LOCKSS (<http://lockss.org>)

Internet Archive (<http://archive.org/>)

LuKII (www.lukii.hu-berlin.de/)

UC3 Merritt (<https://merritt.cdlib.org/>)

MetaArchive Cooperative (<http://www.metaarchive.org>)

Thank you!
Obrigado!

sam@educopia.org
@samalanmeister