Preservation Management in Practice : Trusted Workflow

Li Chunwang, Zhang Xiaolin, Wu Zhenxin

National Science Library, Chinese Academy of Sciences

Outline

1 Introduction to Workflow Management

2 Workflow Management Framework in NSL

3 Workflow Management in Action

4 Further Questions

1 Introduction to Workflow Management

1.1 **Definition**

- “the automation of a business process, in whole or part, during which documents, information or tasks are passed from one participant to another for action, according to a set of procedural rules".

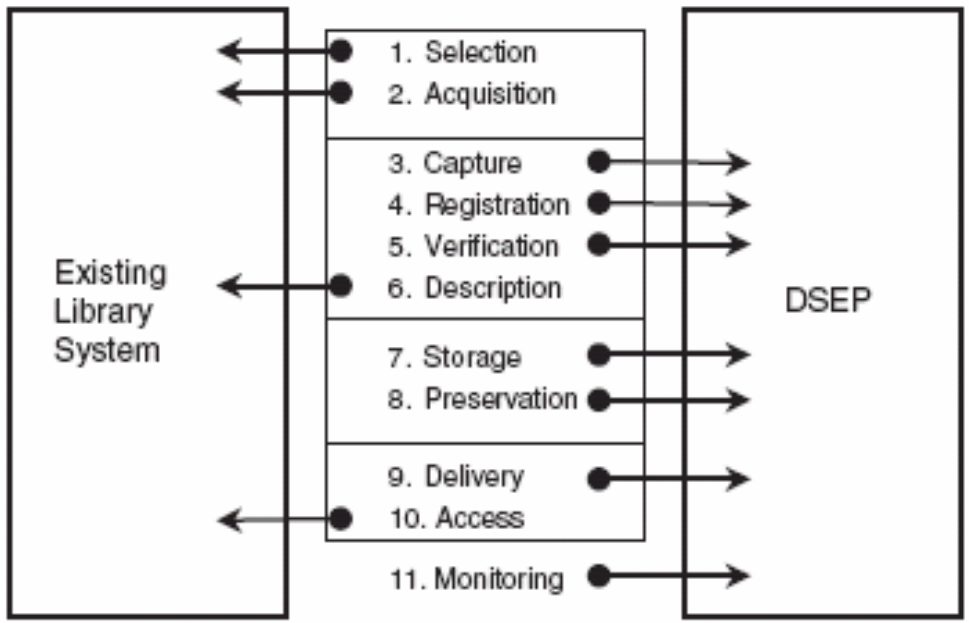
  (Workflow Management Coalition, WfMC. http://www.wfmc.org/)

- “used to describe the tasks, procedural steps, organizations or people involved, required input and output information, and tools needed for each step in a business process.” (http://searchcio.techtarget.com/sDefinition/0,,sid19_gci213384,00.html)

1 Introduction to Workflow Management

1.2 Related Works

- **DSEP Process Model**



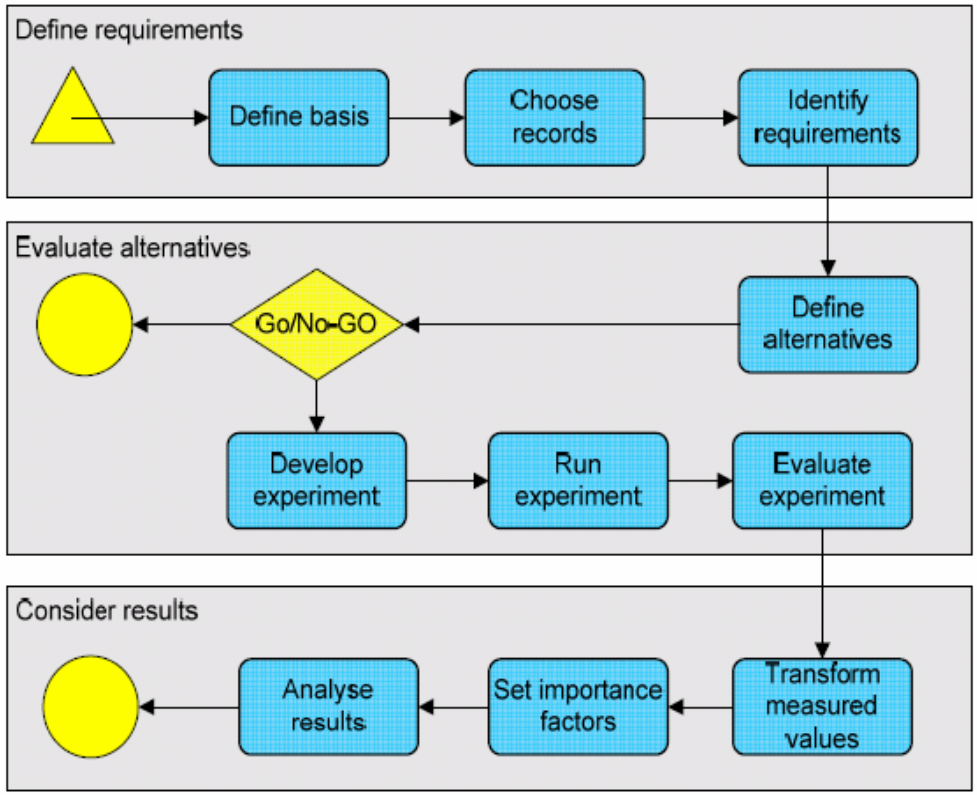
——Titia van der Werf. DESP: A process model. 2000.

1 Introduction to Workflow Management

1.2 Related Works

- **Vienna University of Technology's Planning Workflow**

following the principles of the PLANETS methodology



— Christoph Becker, Proceedings of the 9th Russian Conference on Digital Libraries RCDL 2007, Pereslavl, Russia, 2007

1 Introduction to Workflow Management

1.2 Related Works

- **Nestor:** http://nestor.sub.uni-goettingen.de/index.php
- **Planets :** http://www.planets-project.eu/
- **REMAP:** http://www.hull.ac.uk/remap/index.html
- **PAWN:**http://narawiki.umiacs.umd.edu/twiki/bin/view/Main/PAWN
- **Stephan Heuscher's opinions**

("Workflows in Digital Preservation." -ERPANET Workshop on Workflow, 2004.)

- …

Mostly focus on designing a workflow and workflow automation

1 Introduction to Workflow Management

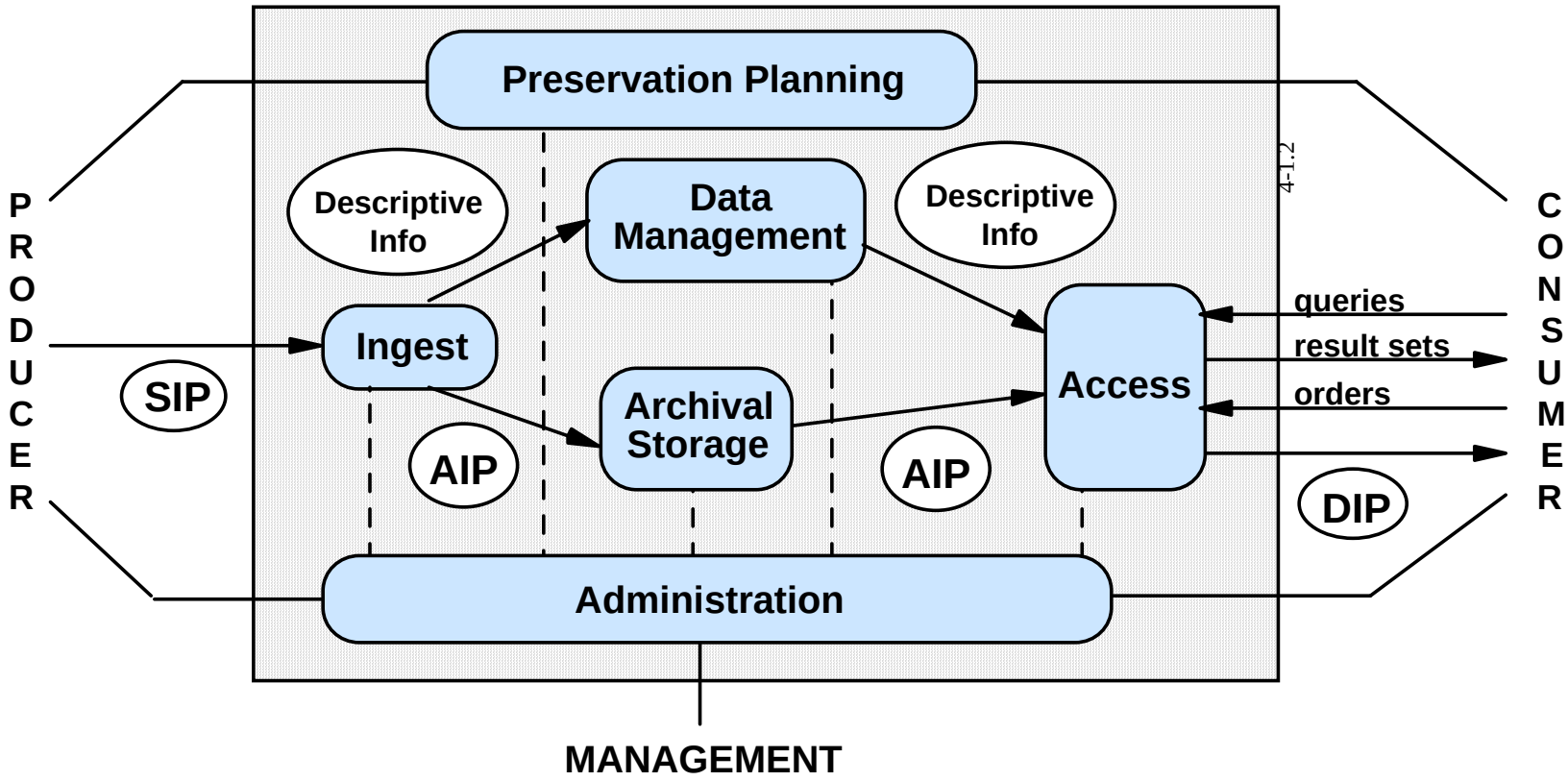
1.3 Why DP need trusted workflow management?

- Expensive
- Complicated
- Uncorrectable immediately
- Multi_participant
- …

2 Workflow Management Framework

2.1 NSL's WfMF(workflow management framework)

- **Based on OAIS**



2 Workflow Management Framework

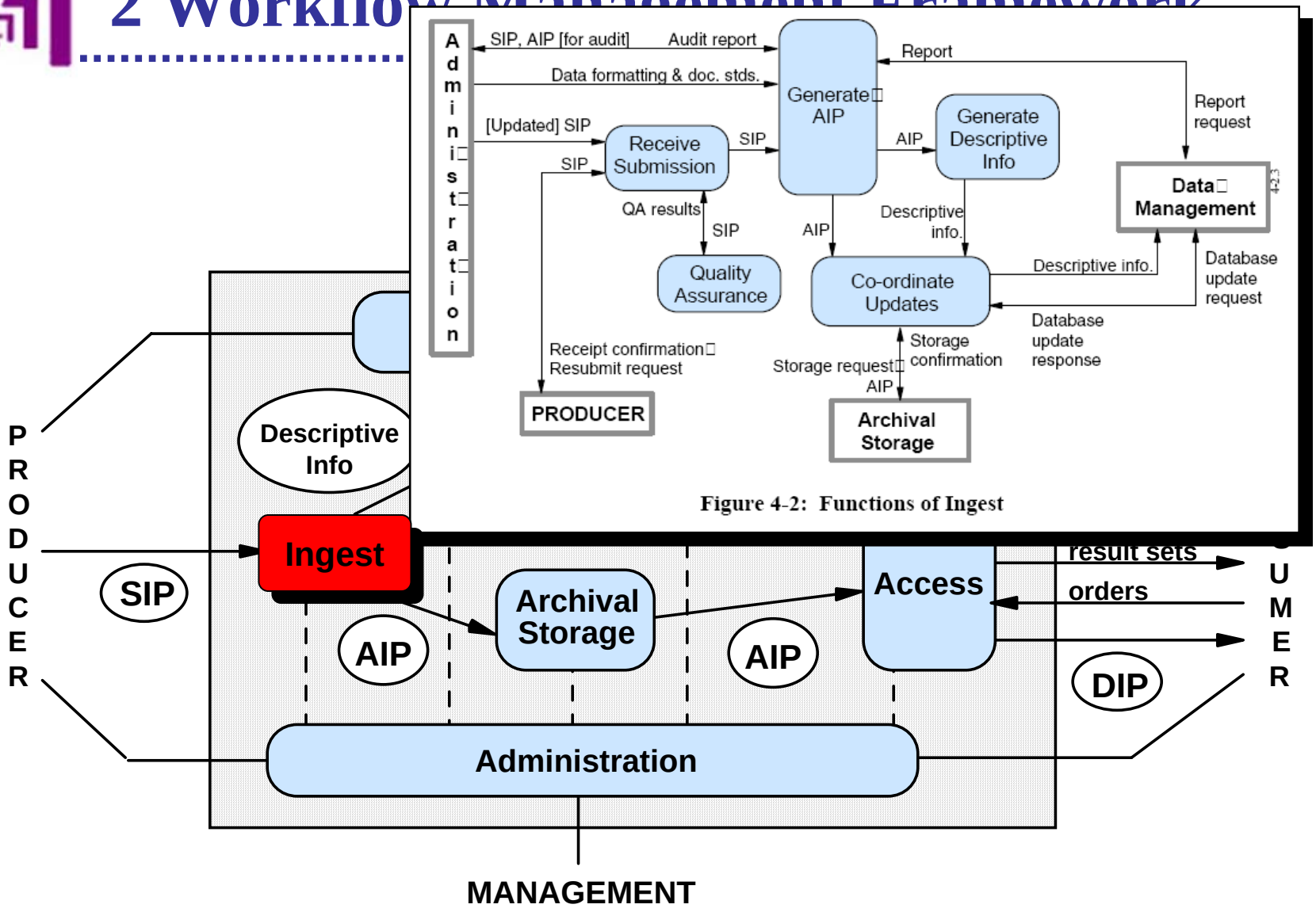
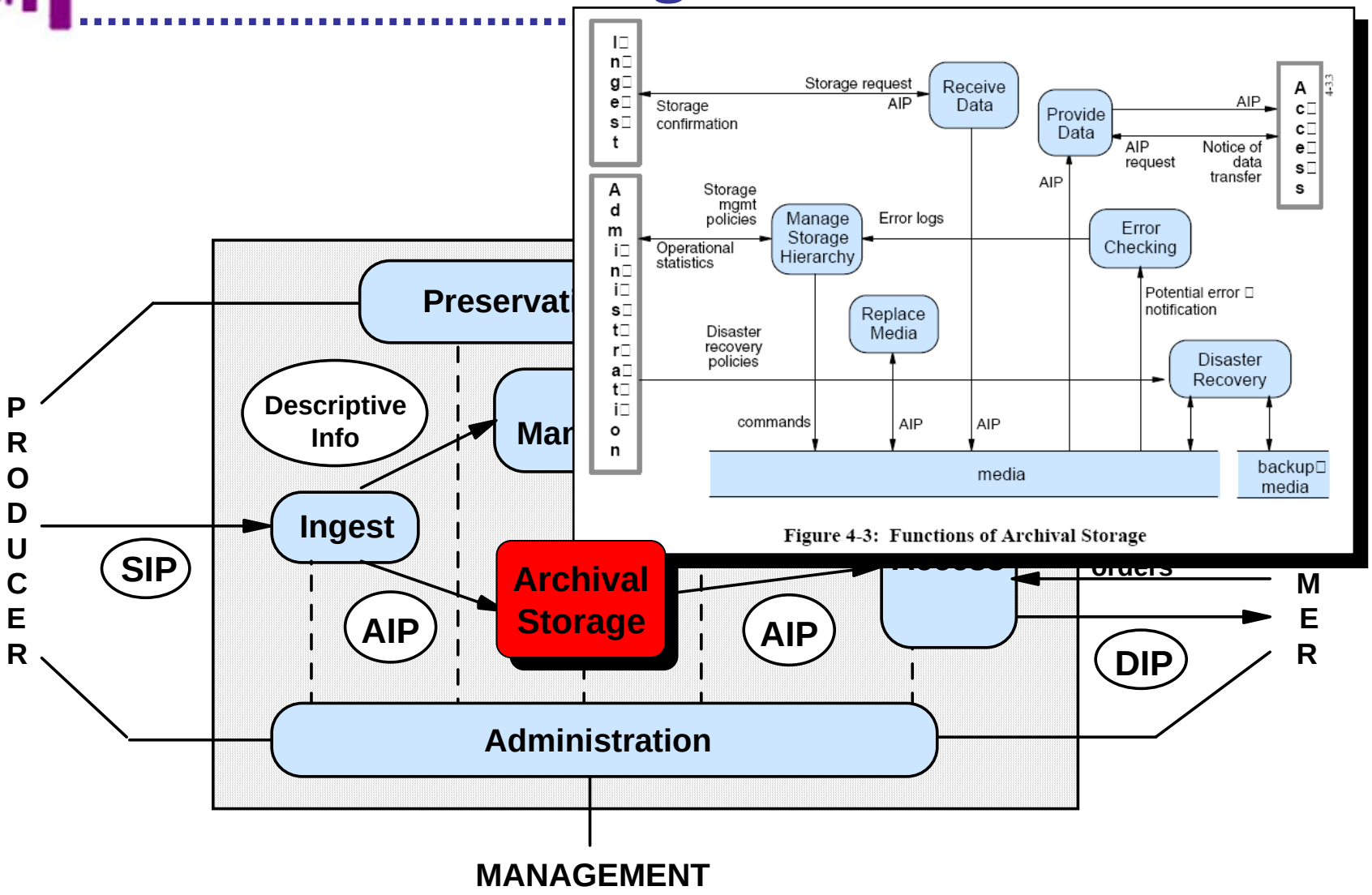
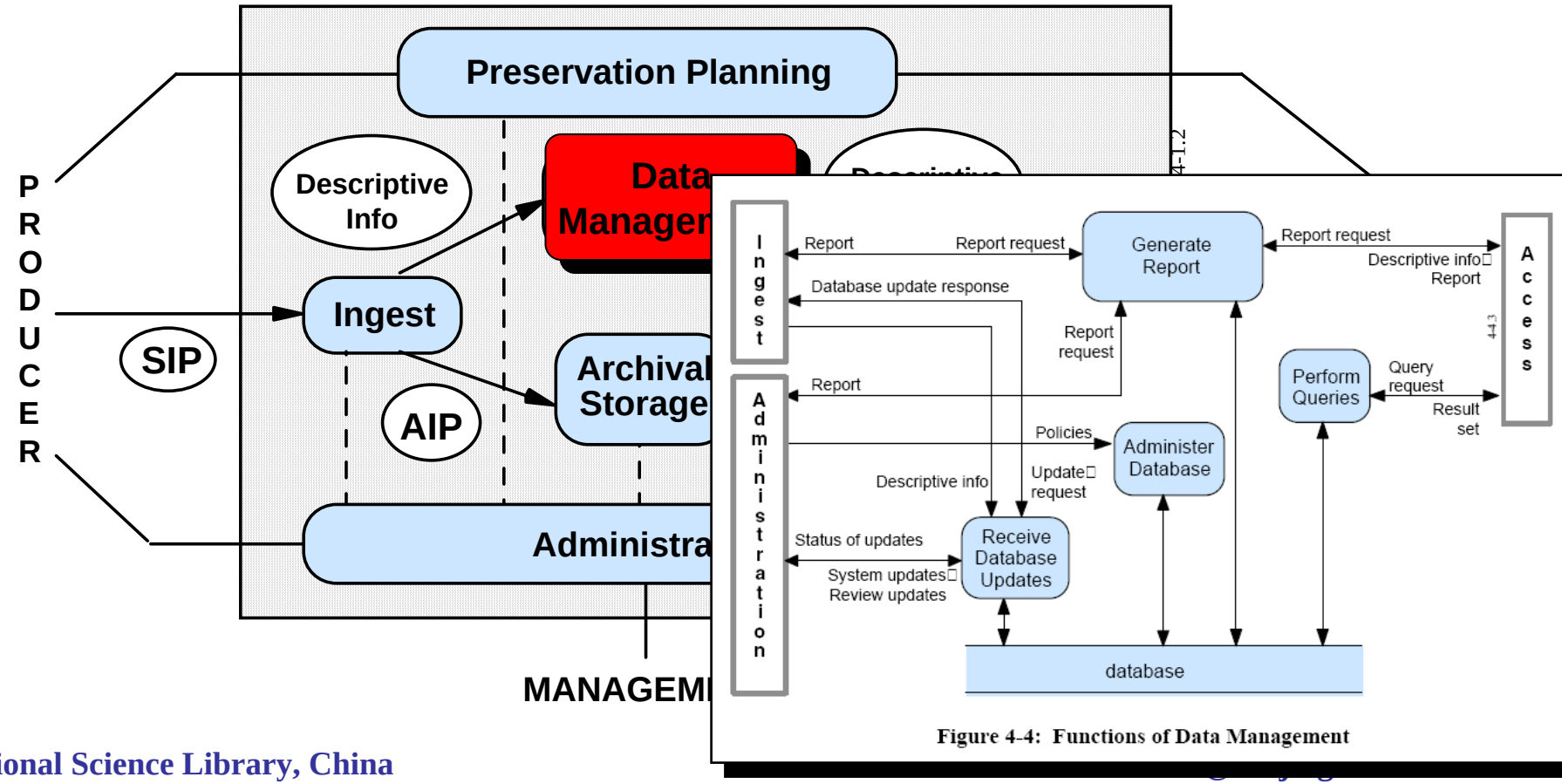


Figure 4-2: Functions of Ingest

2 Workflow Management Framework



2 Workflow Management Framework



2 Workflow Management Framework

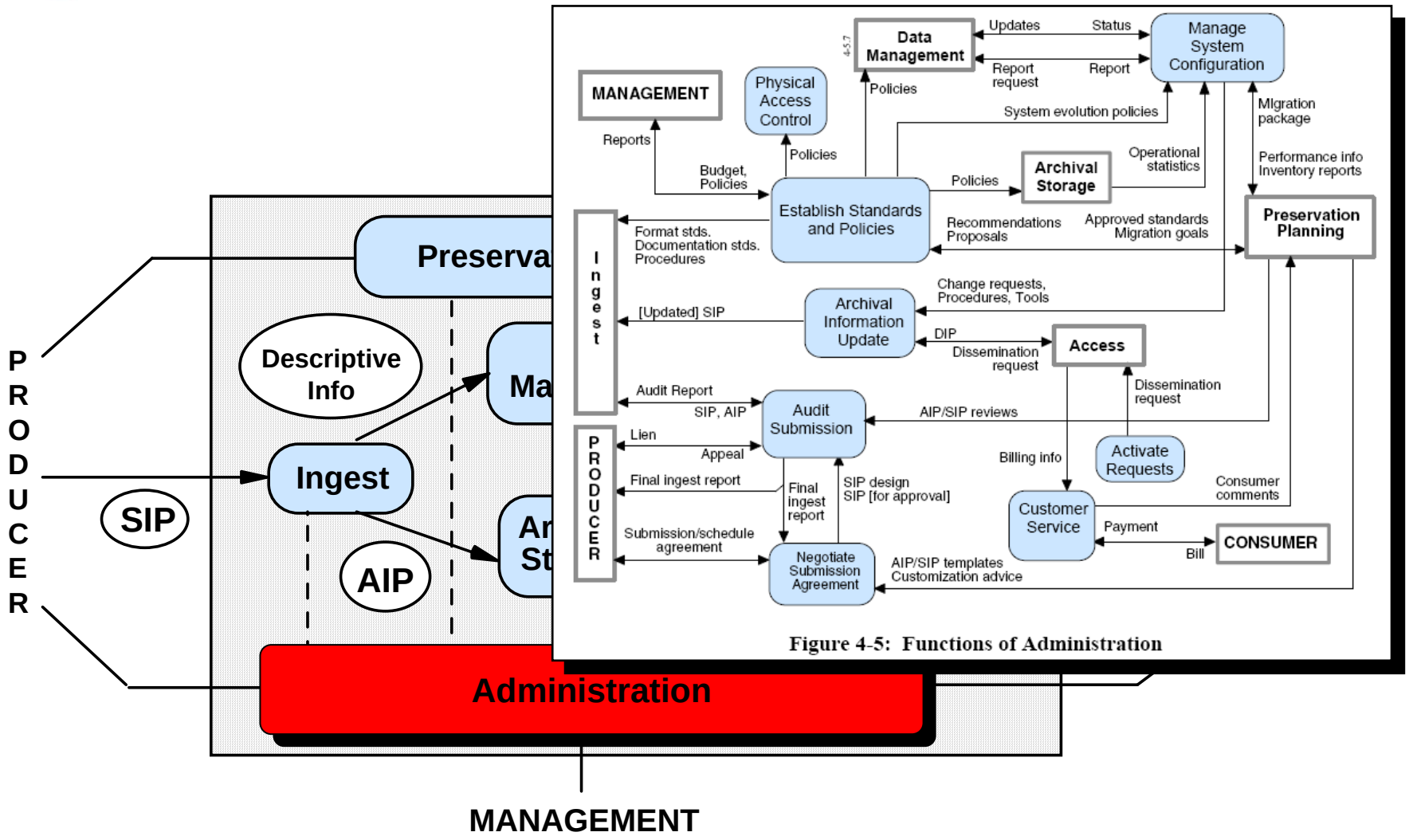


Figure 4-5: Functions of Administration

2 Workflow Management Framework

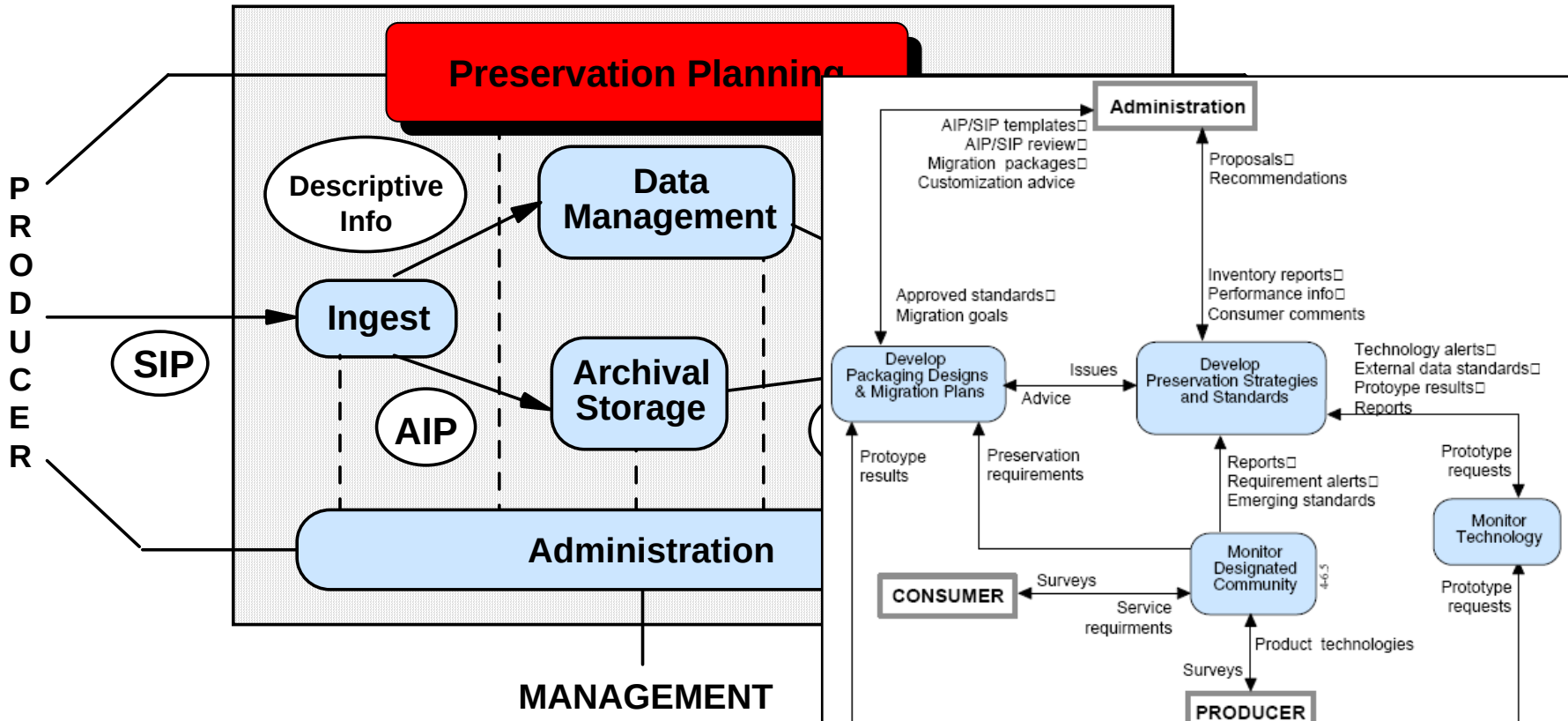
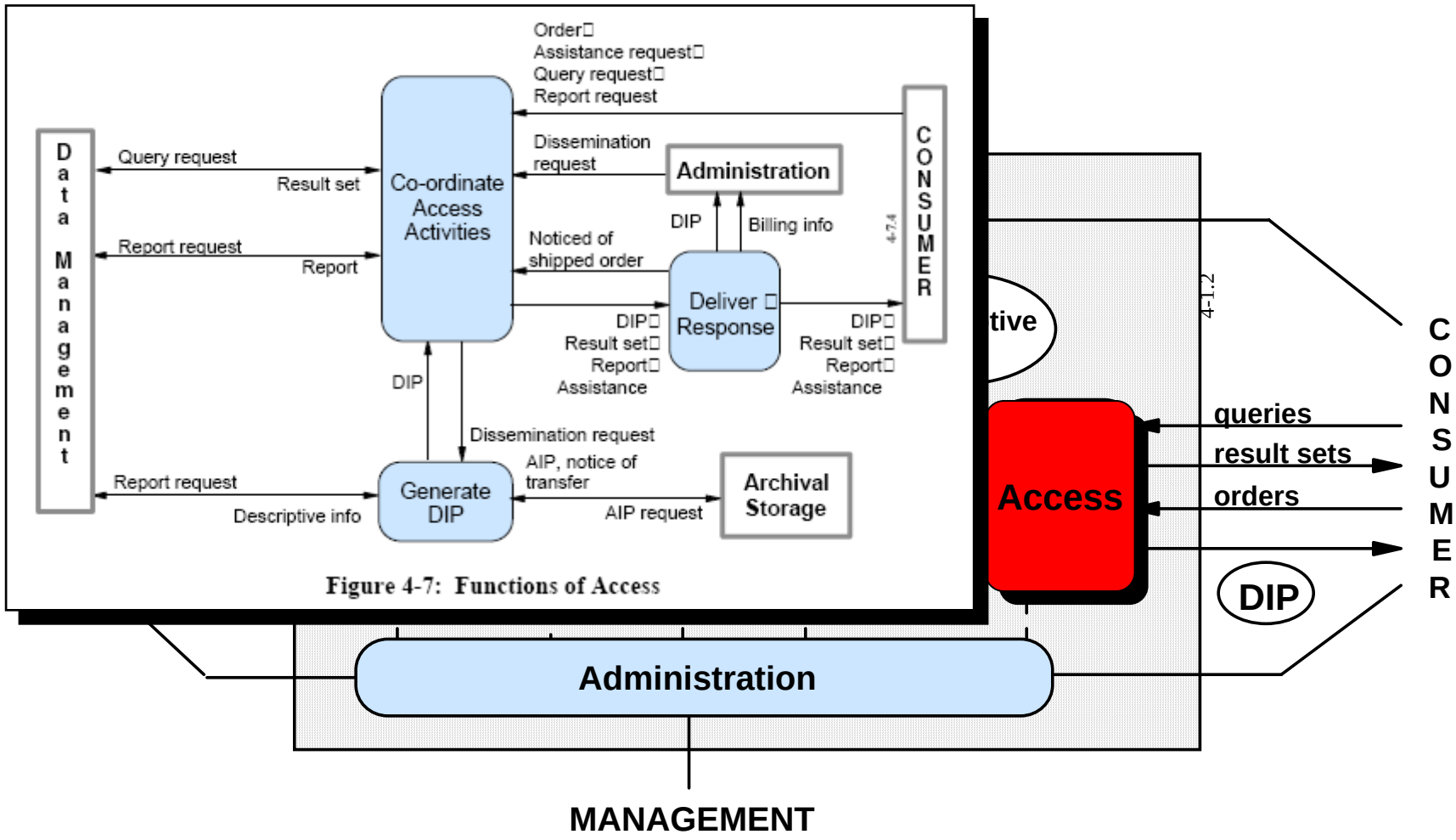


Figure 4-6: Functions of Preservation Planning

2 Workflow Management Framework



2 Workflow Management Framework

2.2 Attributes of trusted workflow management

- Clear defined processes and relations
- Well documents
- Clear responsibility
- With necessary resource support(tools,methods and funds,etc.)
- Correcting mistake ability
- Verifiable
- Auditable
- …

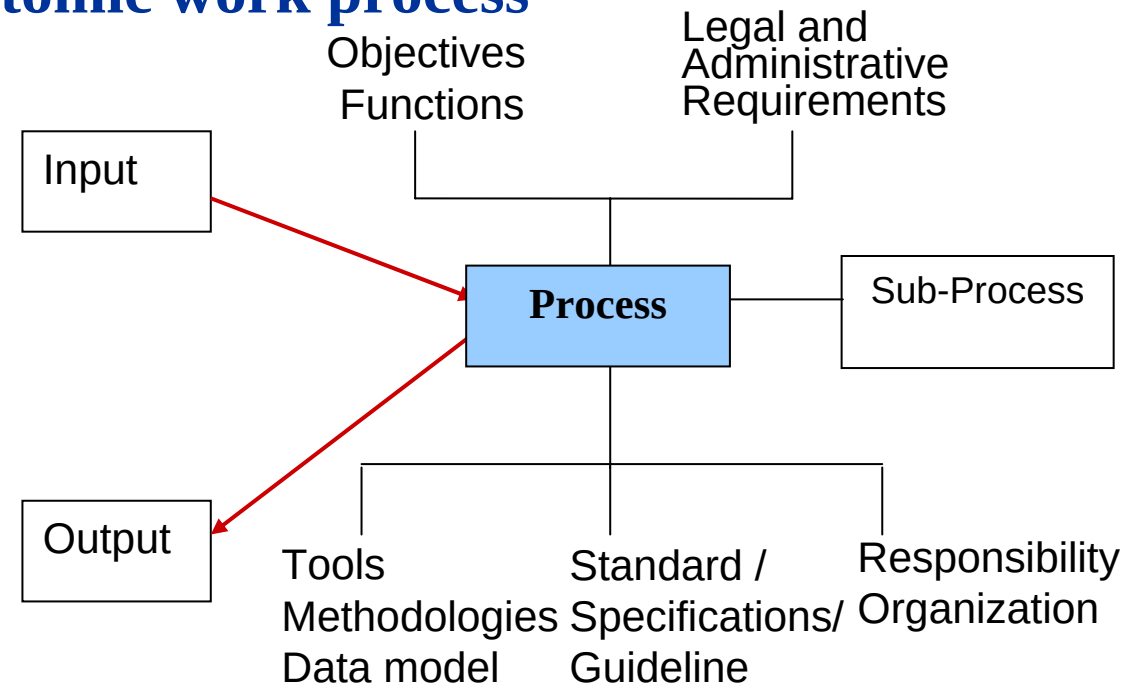
2 Workflow Management Framework

2.3 Atomic work process management model

- **A workflow makes up of processes,and a process makes up of atomic work process**

What, in what format, how much and often, what/how to control, what possible problems

What, in what format, how much and often, what/how to control, what possible problems, to what process



3 Workflow Management in Action

- **In NSL**

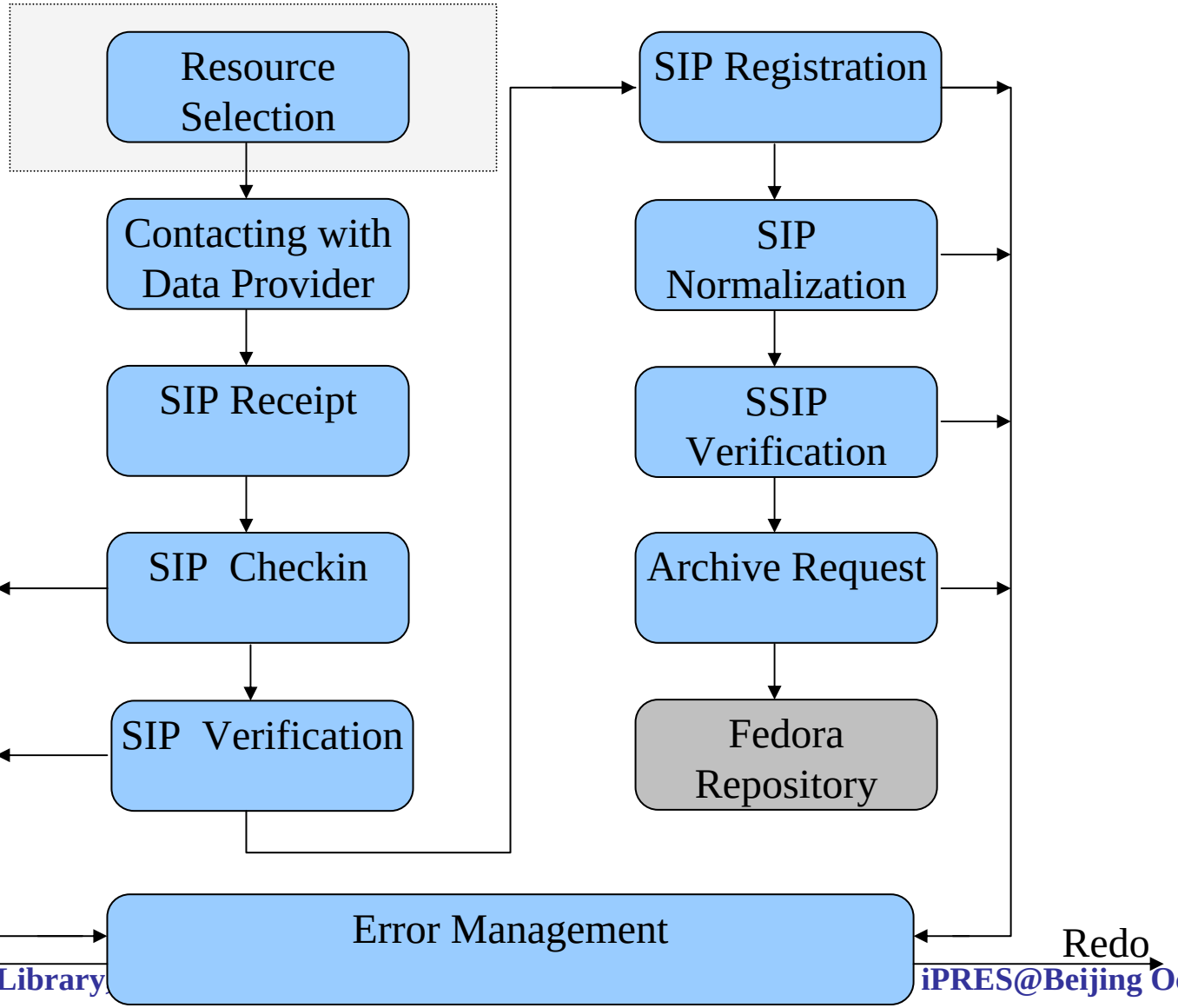
  Developing a CAS Archiving System as part of NSTL Archiving Network, including:

  - Planning and policies
  - An e-journal archiving system
  - Certification of digital repositories
  - Trusted workflow management
  - Others related works

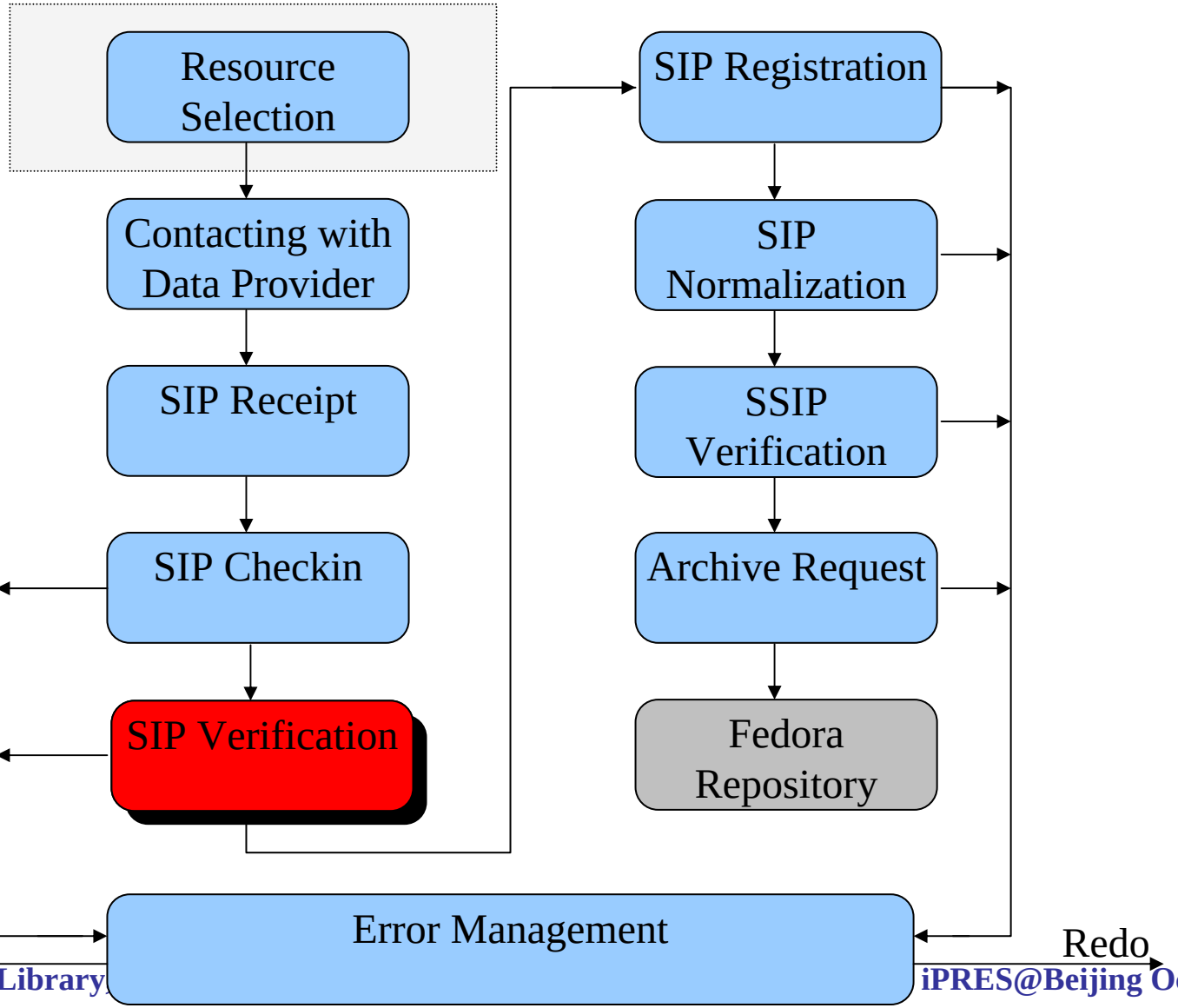
3 Workflow Management in Action

- **In trusted workflow management part of the project:**
  - **Scan this field for knowledge and experiences**
  - **Research the major problems for application**
  - **Develop a trusted workflow management framework**
  - **Develop a trusted workflow management guideline**
  - **Test**

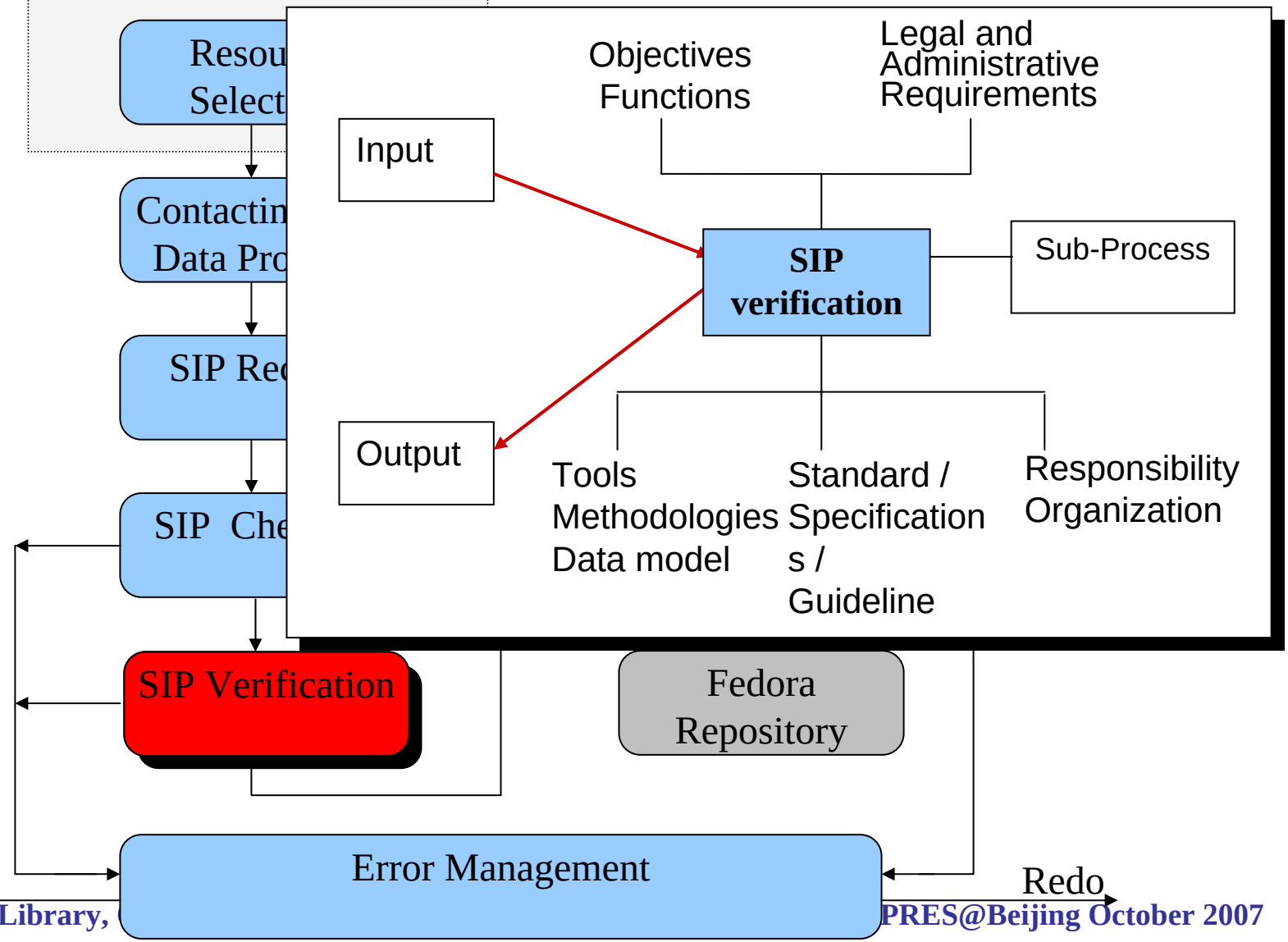
3 A Case Study : Ingest



3 A Case Study : SIP verification

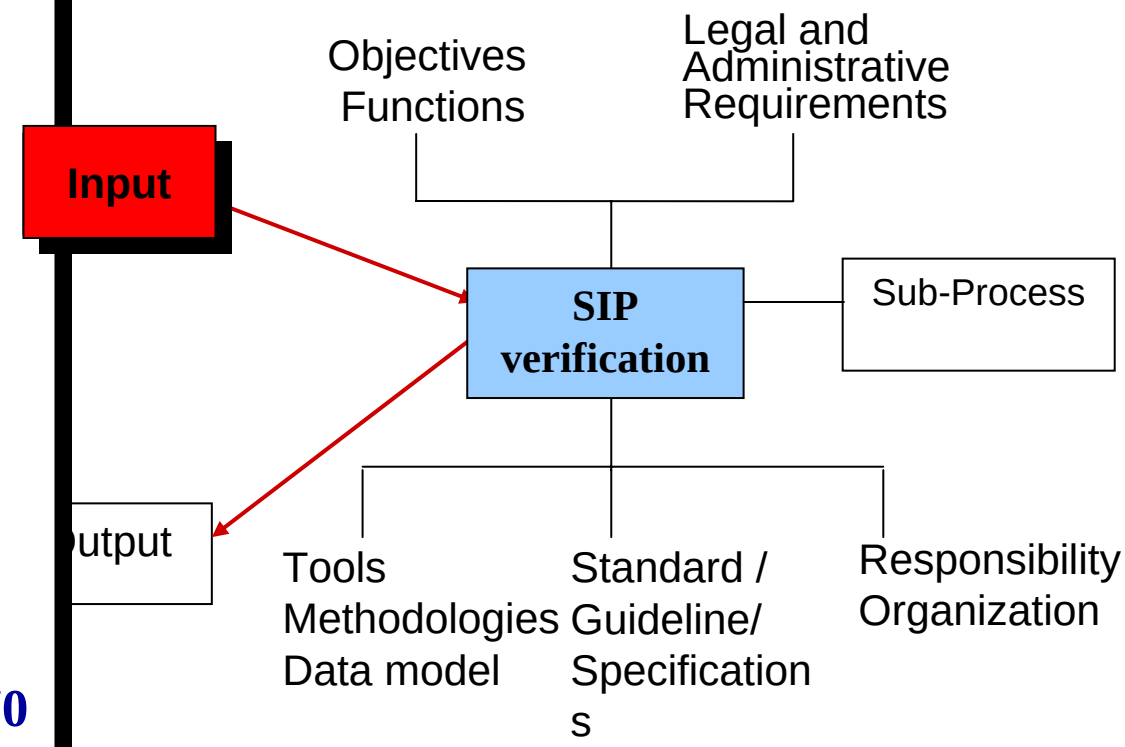


3 A Case Study : SIP verification



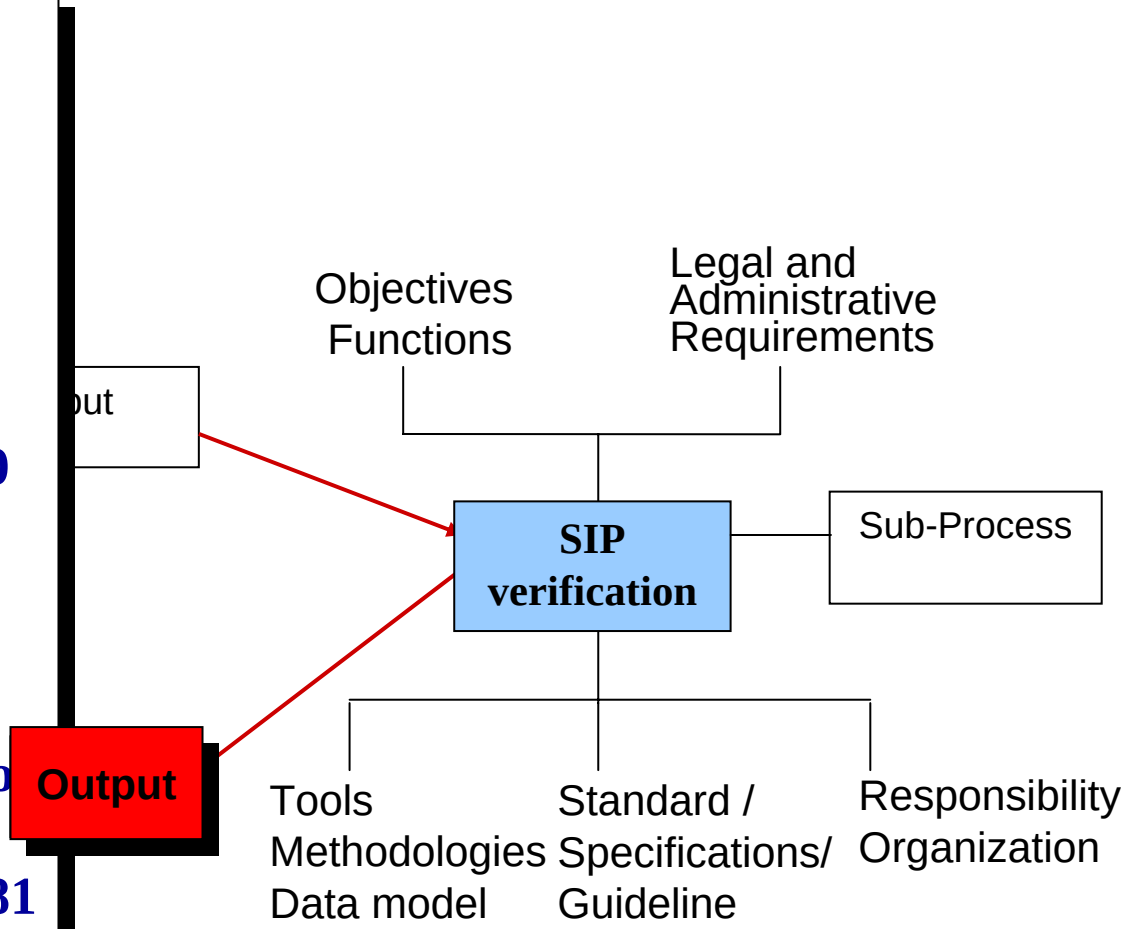
3 A Case Study : SIP verification

Object:SIP from Nature
Source: SIP checkin
Format: XML &PDF
Content:package, dir. , rec.
Submit Req.:
-Frequency: one month
-Coverage: ten journals packages,…
-Time: 2007-08-10 09:53:42.0
-Submitter: Wu Zhenxin
-Method: Request
-Port: None
Description: passed checkin without any error
Related Links: /report/nature/nature20070810_checkin.doc…
Result : Yes(/No/Error)



3 A Case Study : SIP verification

Object:SIP from Data provider
Source: SIP early check
Format: XML &PDF
Content:package, dir. , rec.
Submit Req.:
-Frequency:one month
-Coverage:journals
-Time:2007-10-10 09:53:42.0
-Submitter:Wu Zhenxin
-Method:Request
-Port: None
Description: passed verification without any erro
Related Links: /report/nature/nature20070810_ verification.doc…
Result :Yes/No/Error

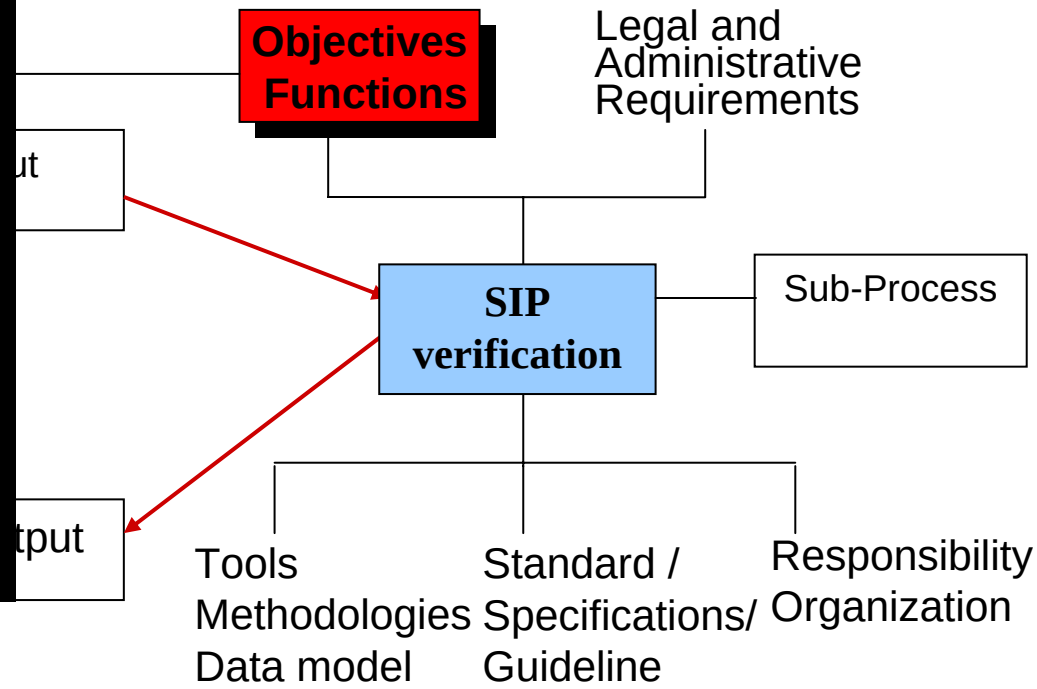


3 A Case Study : SIP verification

- **Objectives & Functions**

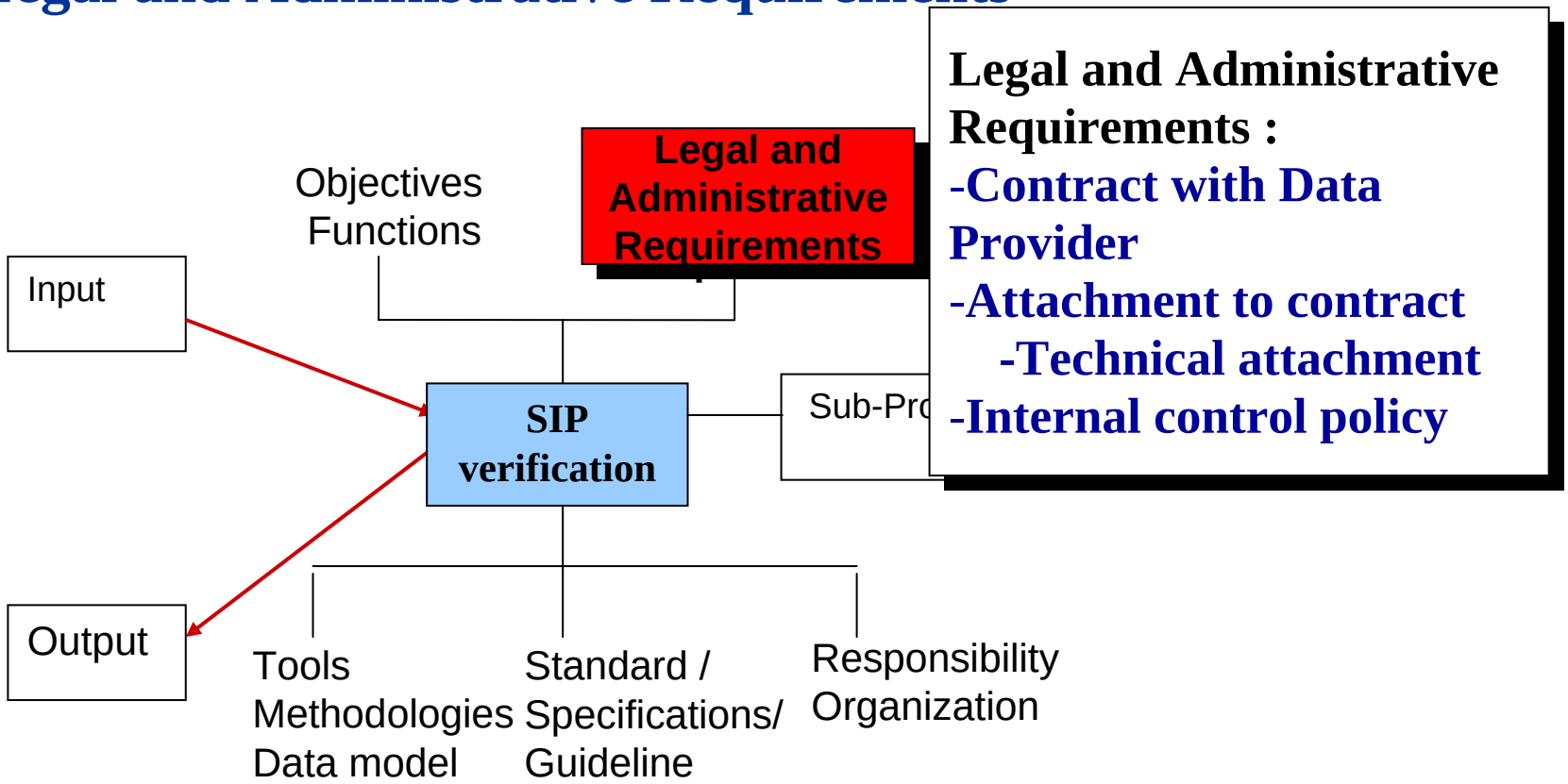
Objectives:
verify SIP according to SA(Submission Agreement)

Functions:
-integrality of all packages
-structure of the XML files
-Version of PDF files
- Number of the package, collection,record



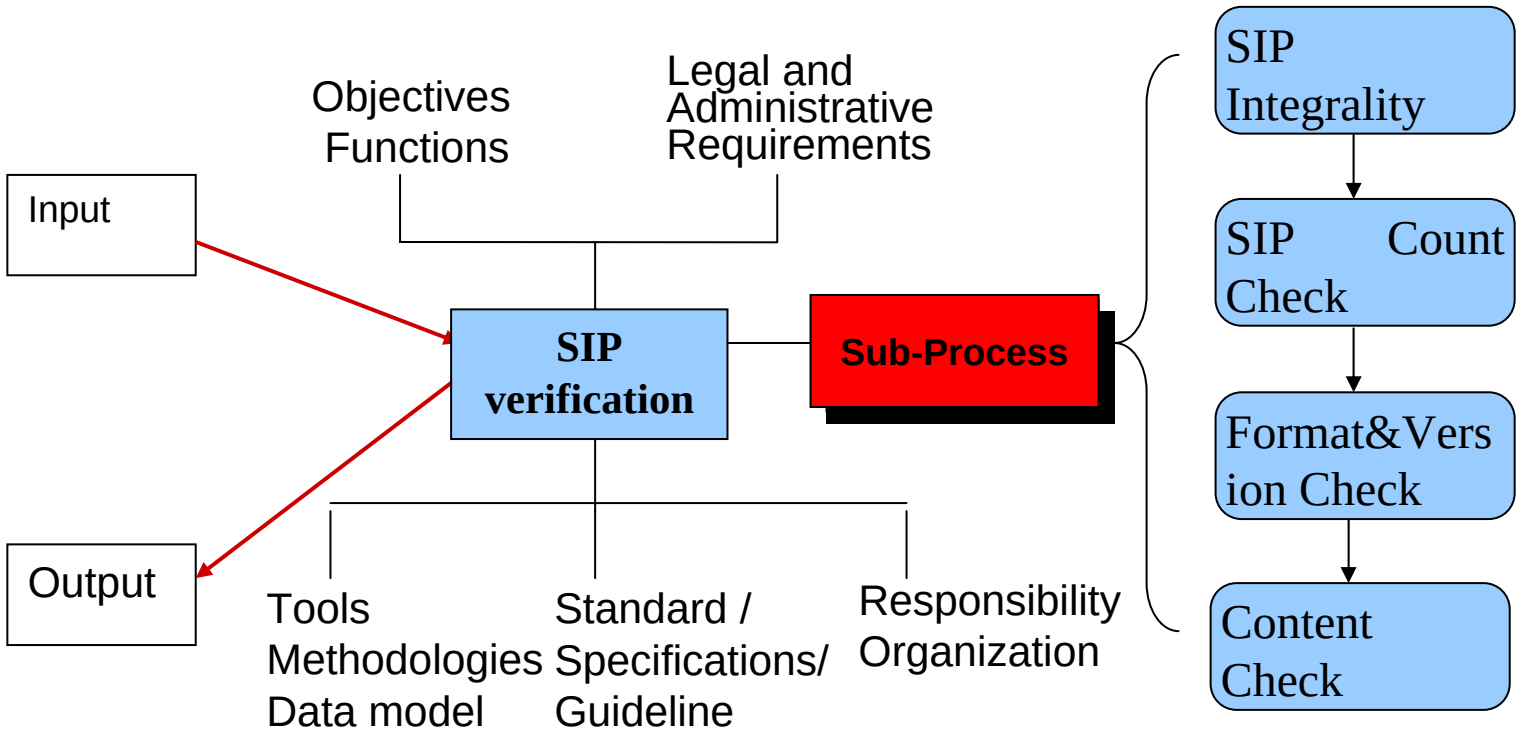
3 A Case Study : SIP verification

- **Legal and Administrative Requirements**



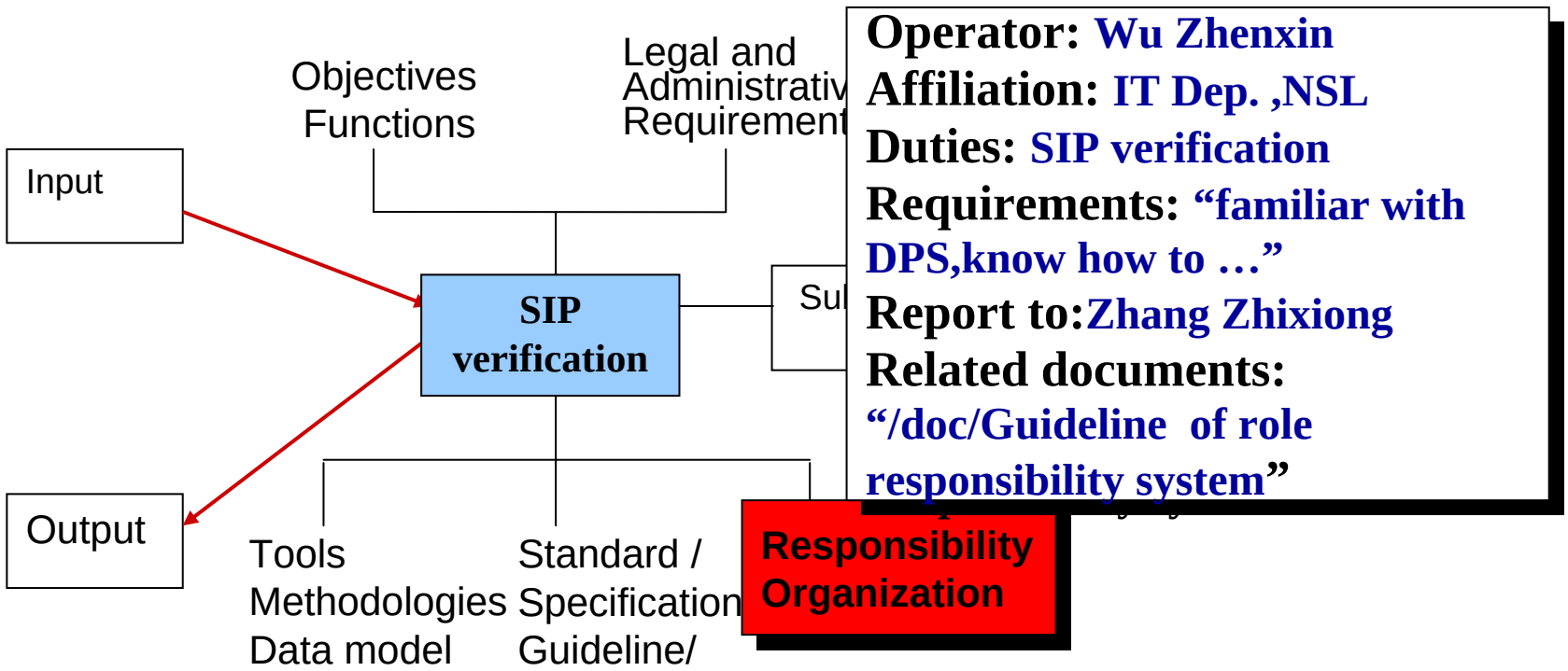
3 A Case Study : SIP verification

- **Sub-Process**



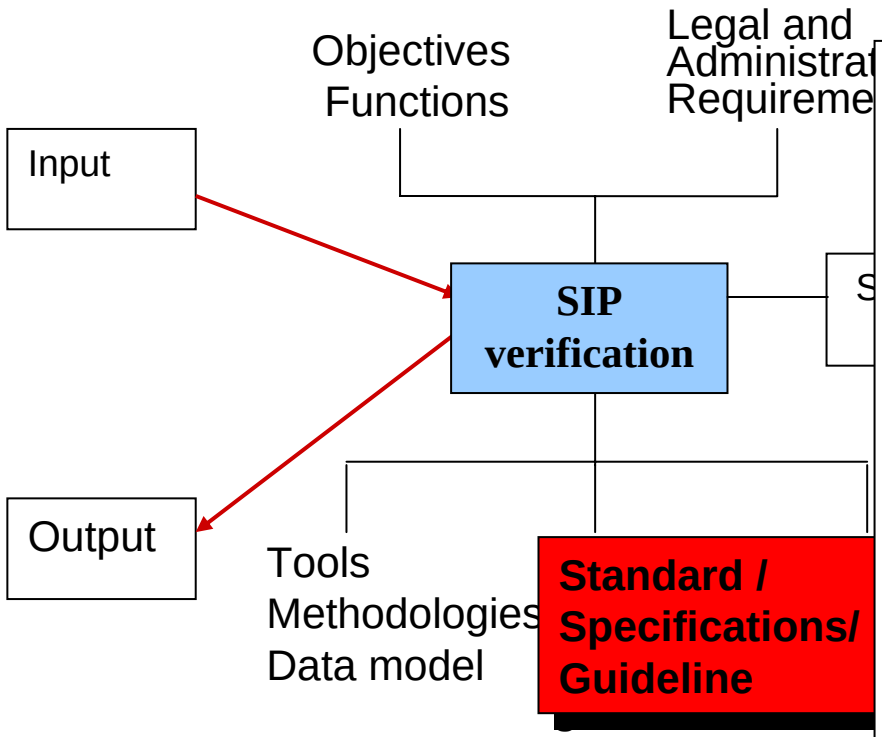
3 A Case Study : SIP verification

- **Responsibility Organization**



3 A Case Study : SIP verification

- **Standard / Specifications / Guideline**



A case of specifications

Title: Nature_XML_content_Spec

Type: XML

Use for:XML content verification

Creator:nature

Issue time:2007-06-10

Uri:http://10.0.11.54:8080/DPS/doc/

Valid date:2008-06-10

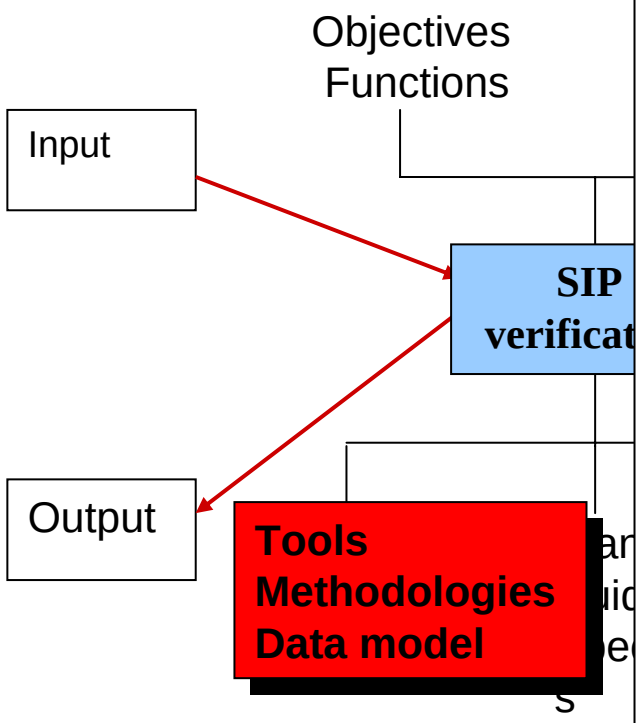
Description:metadata file format and structure description

Related documents:

Nature_tech_attachment

3 A Case Study : SIP verification

- **Tools / Methodologies / Data model**



A case of tools

Title:XML_content_verification
Type: inner_program
Use for:XML content verification
Creator: Liu Jianhua, IT dep.
Issue time:2007-08-10
Uri:http://10.0.11.54:8080/DPS/class/
Valid date:2008-08-10
Description:analyze metadata file structure and content according by tech attachment from data provider…
Related documents:
Nature_tech_attachment
Springer_tech_attachment…

4 Further questions

- Enforceable granularity
- Labor intensive
- It’s very burdensome
- Implementation
- Collaboration with other parts
- Best practice
- …

- **Thanks my colleague for many helps**
  - **Zhang Zhixiong**
  - **Huang Guobin**

Question?

Wu Zhenxin

Wuzx@mail.las.ac.cn

Thanks !