# Preservation Planning in the OAIS Model

Stephan Strodl, Andreas Rauber
Vienna University of Technology, Vienna, Austria

IPRES 2007

Beijing, China 11-12. October 2007

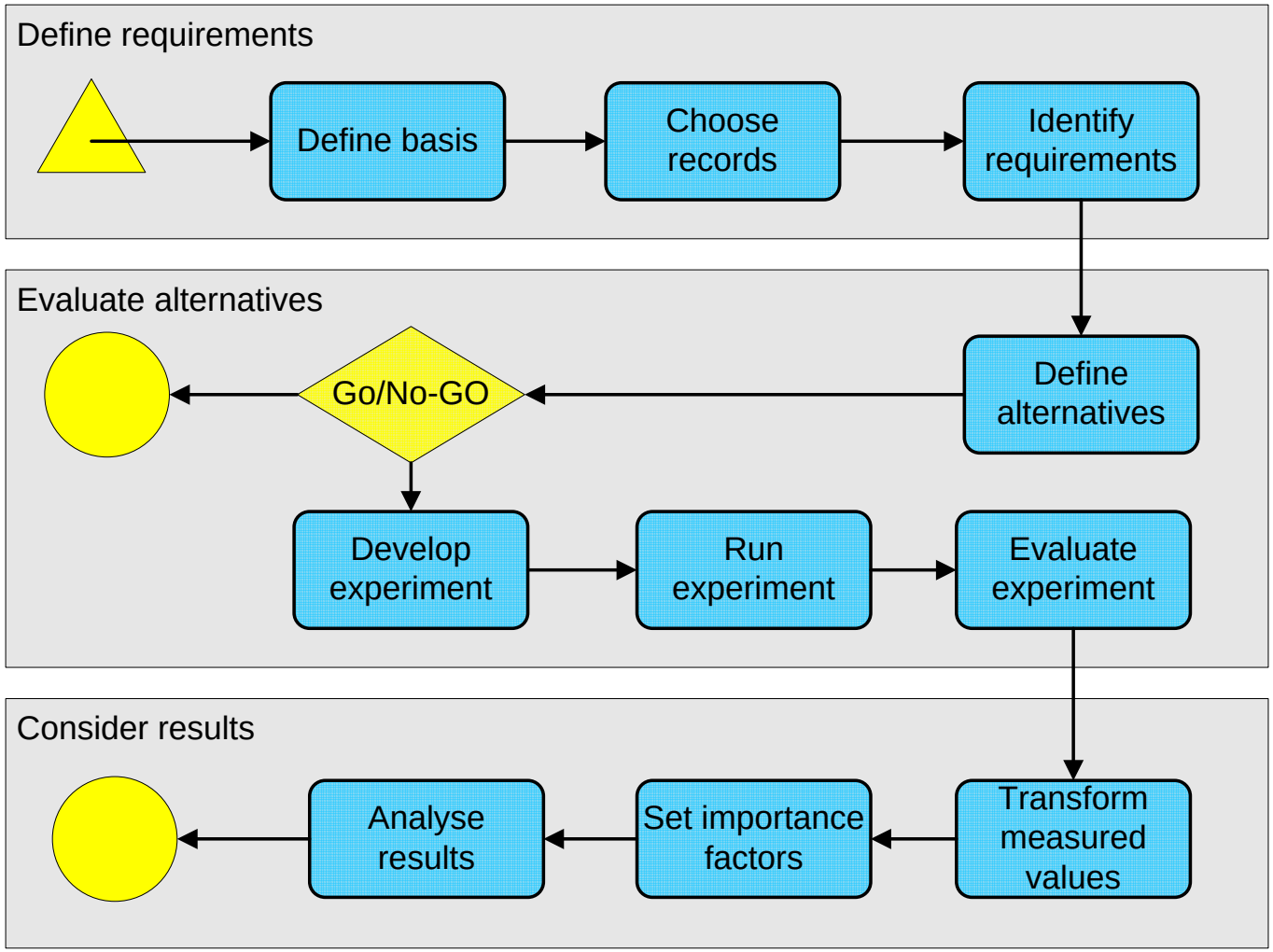
# Overview

- Introduction
  - Planets Preservation Planning approach
- Preservation Planning in the OAIS Model
  - Alerts
  - Preservation Planning Information Flow
  - Effects of new Preservaiton Strategies
- Conclusions

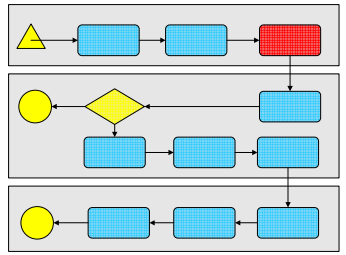
# Introduction

- OAIS model
  - Wide acceptance
  - Design reference model for an archival system
  - Without detailed activity requirements
- Preservation Planning
  - Different strategies
  - Different tools for each strategy
  - Different parameters for each tool
  - Which solution to select to meet preservation requirements

# Preservation Planning approach

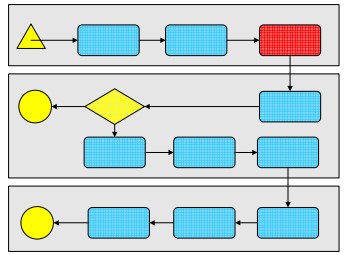


# Objective Tree



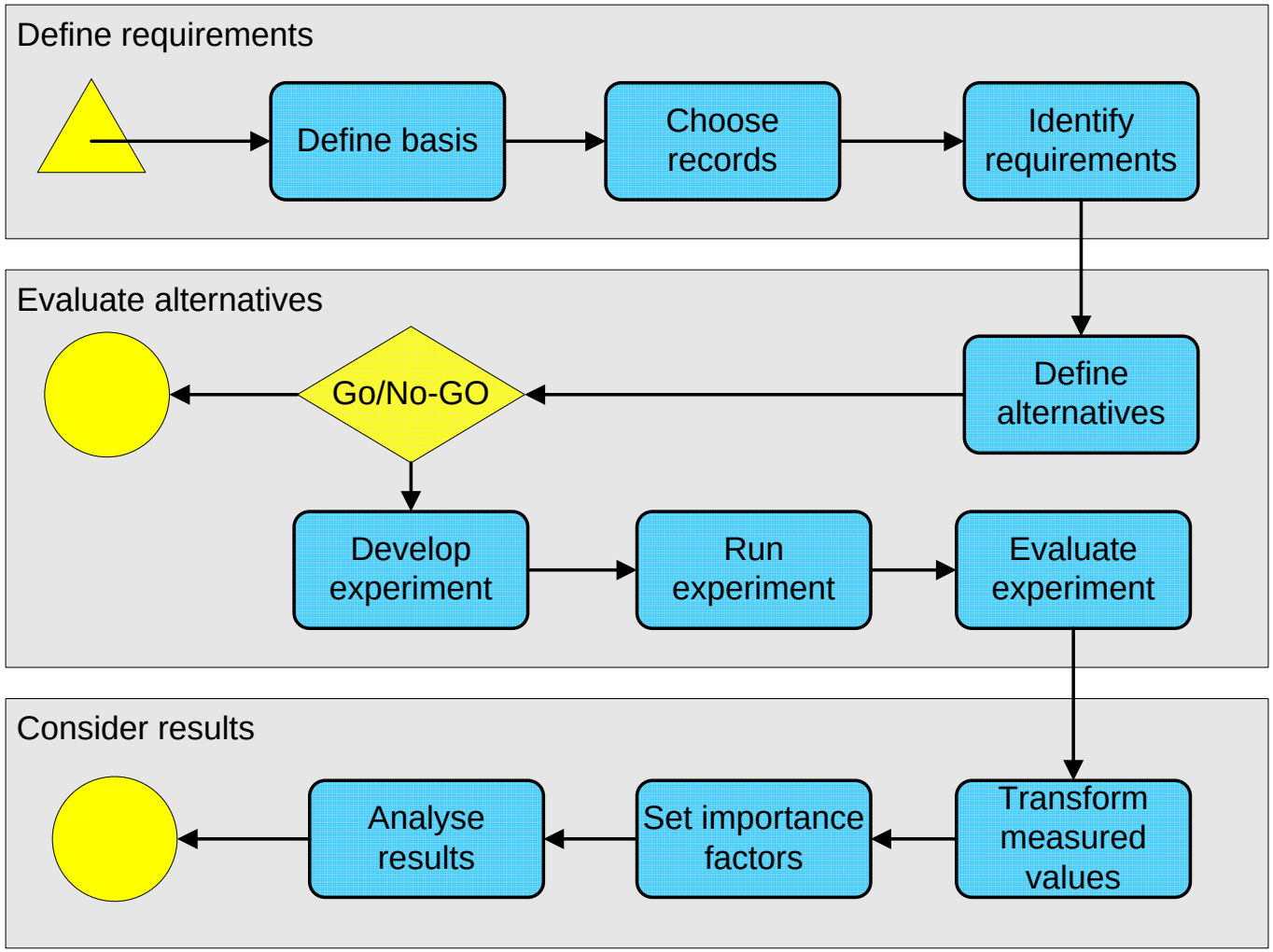
- Website
  - Record characteristics
    - Technical characteristics
      - Ubiquity: Ubiquitous/Widespread/Specialised/Obsolete
      - Support: 0/1-5/6-10/10+
      - Documentation
        - Quality: Primary/Secondary
        - Disclosure: Full/Partial/None
        - Openness: Standard/Open/Proprietary
        - Availability: Public/Limited/None
      - Stability
        - Speed of change: <1 year/1-2 years/3-5 years/>5 years
        - Backwards compatability: None/Previous version only/ Some previous versions/All previous versions
      - Ease of identification: Automatic/Manual/No
      - Ease of validation: Automatic/Manual/No
      - Lossiness: Lossy/Lossless
      - IPR: Y/N
      - Complexity: High/Medium/Low
      - Error tolerance: None/Detectable/Recoverable
      - Comparative size: Larger/Same/Smaller
    - Infrastructure characteristics
      - Hardware
        - Acquisition: £
        - Maintenance: £/annum
      - Software
        - Acquisition: £
        - Maintenance: £/annum
      - Staff
        - Availability: Widespread/Scarce/Not available
        - Training
          - Cost: £/person
        - Operations
          - Numbers: Actual numeric value
          - Cost: £/person/annum
    - Process characteristics
      - Automation: Pre-configured/API/Scriptable/None
      - Performance
        - Streamable: Y/N
        - Throughput: Files/Hour
        - Scalability: Y/N
      - Documentation
        - Format: Compatible/Transformable/Incompatible
        - Content: Comprehensive/Satisfactory/Minimal/None
      - Useability: Good/Acceptable/Poor
      - Validation
        - Syntactic: Full/Partial/None
        - Semantic: Full/Partial/None
    - Appearance
      - Fonts: Y/N
      - Page layout: Y/A/N
      - Highlighting
        - Links: Y/N
        - Text: Y/N
        - Cells: Y/N
    - Content
      - Documents
        - Text content: Y/N
        - Encoding: Text/Image
      - Images
        - Width: Y/N
        - Height: Y/N
        - Resolution: Y/N
        - Bit depth: >= original/< original
        - Colour space: Y/N
        - histogram: ?
      - Internal links present and working: Y/N
      - Tables
        - rows unchanged
        - columns unchanged
        - cell content unchanged
    - Structure
      - Pagination
        - No. of pages: Y/N
        - Page breaks: Y/A/N
      - Document structure
        - Heading structure: Y/N
        - Headers: Y/N
        - Footers: Y/N
        - TOC: Y/N
        - Appendices: Y/N
        - Sequence: Y/N
    - Behaviour
      - deactivate
        - mailto: y/N
      - preserve
        - menus: complete/navigable/missing
        - pop-ups: Y/N
      - freeze
        - current date/time: frozen/missing/current
        - visitor counter: frozen/missing/current
        - Newsfeeds: frozen/missing/current
    - Context
      - External links present and unchanged: Y/N
      - Links within the collection present and valid: Y/N

# Objective Tree



- Technical characteristics
  - Stability
    - Backwards compatability: None/Previous version only/ Some previous versions/All previous versions
    - Speed of change: <1 year/1-2 years/3-5 years/>5 years
  - Documentation
    - Availability: Public/Limited/None
    - Openness: Standard/Open/Proprietary
    - Disclosure: Full/Partial/None
    - Quality: Primary/Secondary
    - Support: 0/1-5/6-10/10+
    - Ubiquity: Ubiquitous/Widespread/Specialised/Obsolete
  - Comparative size: Larger/Same/Smaller
  - Error tolerance: None/Detectable/Recoverable
  - Complexity: High/Medium/Low
  - IPR: Y/N
  - Lossiness: Lossy/Lossless
  - Ease of validation: Automatic/Manual/No
  - Ease of identification: Automatic/Manual/No

# Preservation Planning approach



# Result of Preservation Planning

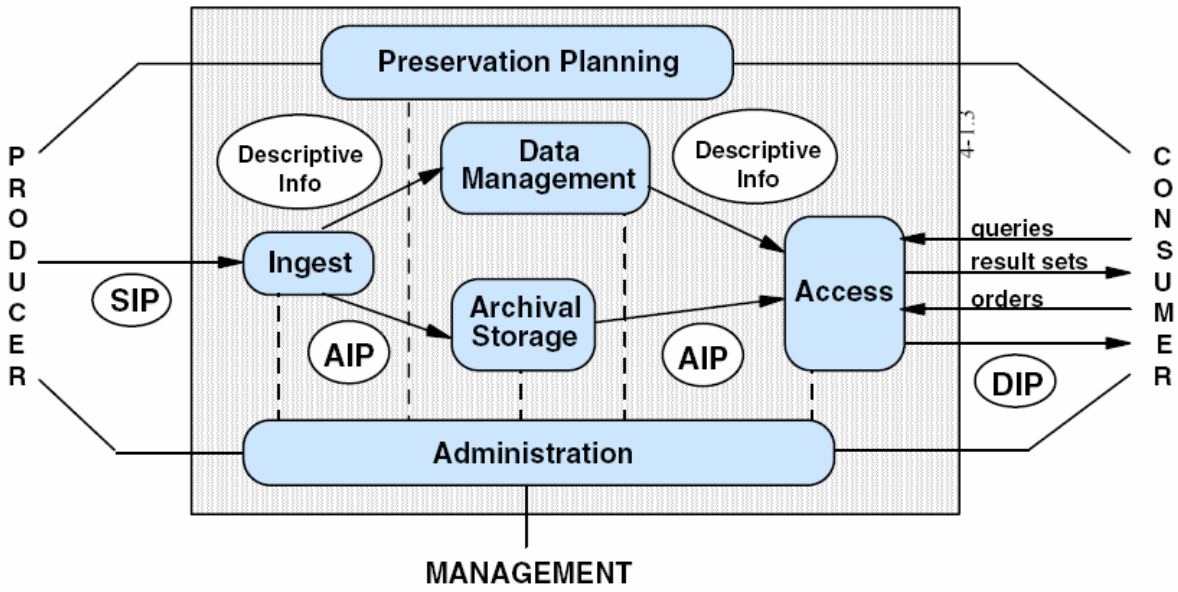
- Austrian National Library
- Master Theses from Austrian Universities

| Nr. | Alternative | Total Score | |
|---|---|---|---|
| | | Sum | Multiplication |
| 1 | PDF/A (Adobe Acrobat 7 prof.) | 4.52 | 4.31 |
| 2 | TIFF (Document Convert 4.1) | 4.26 | 3.96 |
| 3 | EPS (Adobe Acrobat 7 prof.) | 4.22 | 3.91 |
| 4 | JPEG (Adobe Acrobat 7 prof.) | 4.17 | 3.77 |
| 5 | RTF (Adobe Acrobat 7 prof.) | 3.43 | 0.0 |
| 6 | RTF (ConvertDoc 4.1) | 3.38 | 0.0 |
| 7 | TXT (Adobe Acrobat 7 prof.) | 3.28 | 0.0 |

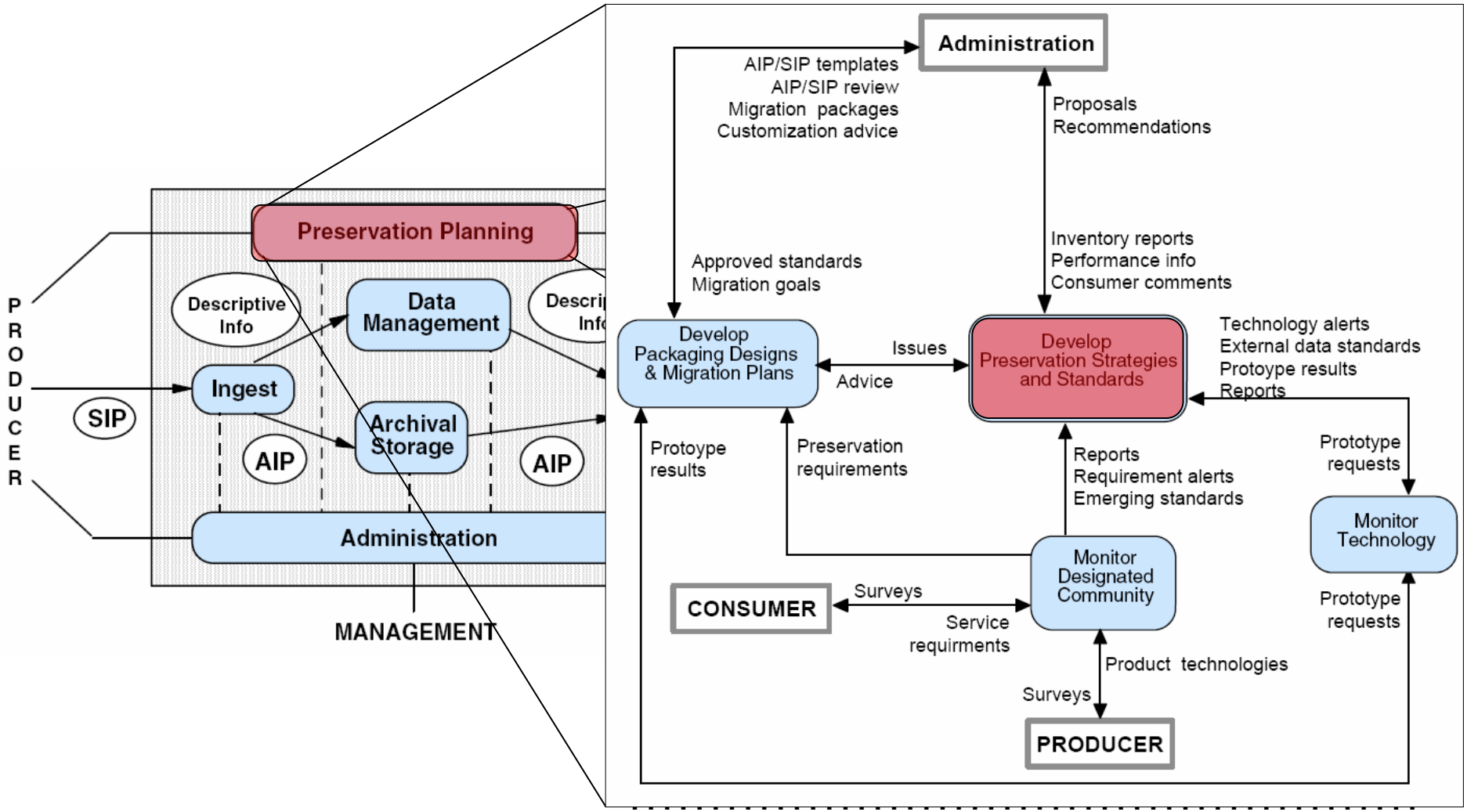
# Preservation Planning approach

- Definition of specific requirements and goals
- Evaluation by experiments on sample records from the collection
- Numeric results based on the Utility Analysis
- Comparison of different preservation strategies

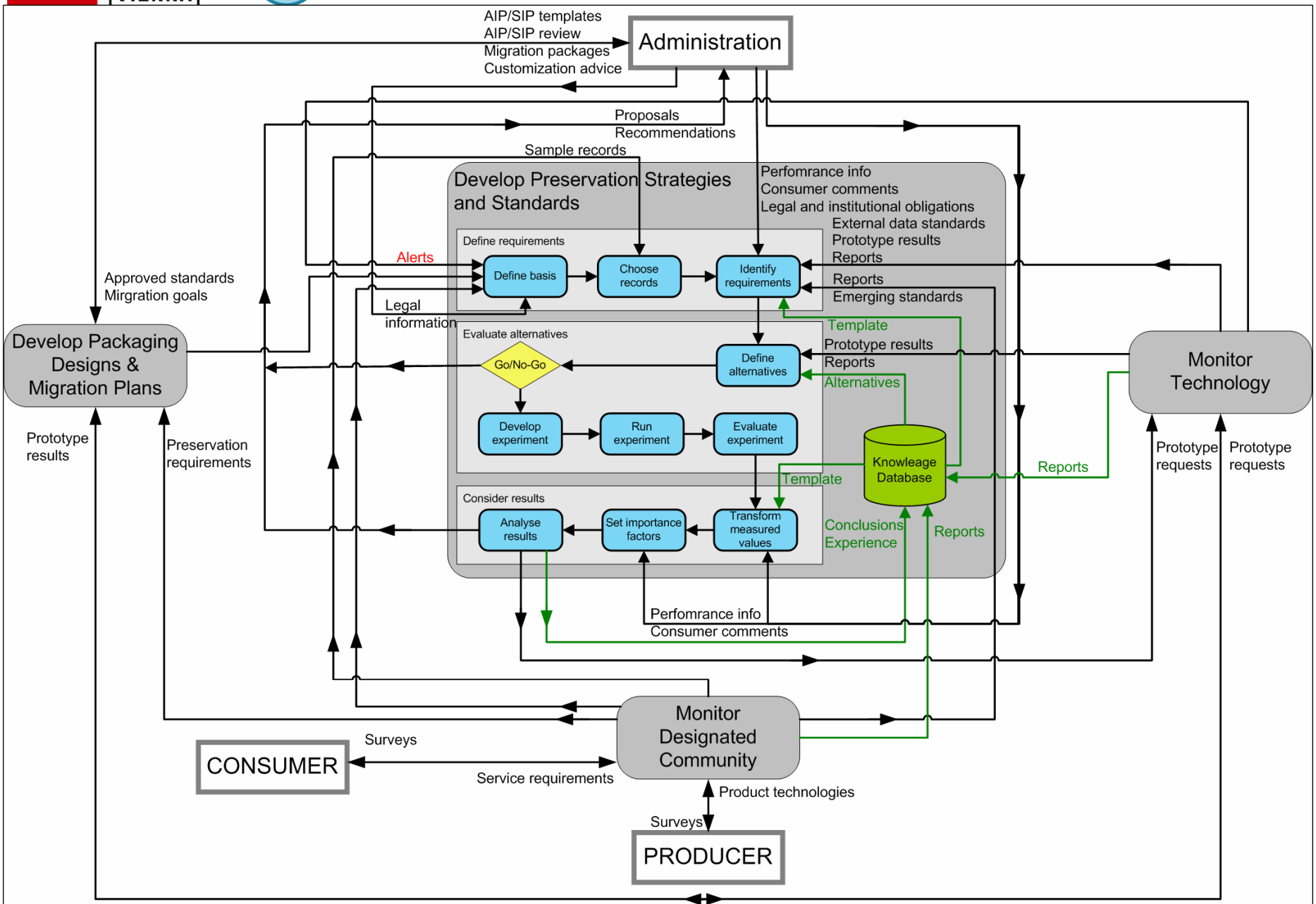
# OAIS Model


Preservation Planning
Descriptive Info
Data Management
Descriptive Info
PRODUCER
Ingest
SIP
AIP
Archival Storage
AIP
Access
queries
result sets
orders
DIP
CONSUMER
Administration
MANAGEMENT
4-1.3

# OAIS Model


PRODUCER
SIP
Preservation Planning
Descriptive Info
Data Management
Ingest
Archival Storage
AIP
AIP
Administration
MANAGEMENT
Administration
AIP/SIP templates
AIP/SIP review
Migration packages
Customization advice
Proposals
Recommendations
Approved standards
Migration goals
Inventory reports
Performance info
Consumer comments
Develop Packaging Designs & Migration Plans
Issues
Advice
Develop Preservation Strategies and Standards
Technology alerts
External data standards
Protoype results
Reports
Protoype results
Preservation requirements
Reports
Requirement alerts
Emerging standards
Prototype requests
Monitor Technology
Monitor Designated Community
CONSUMER
Surveys
Service requirments
Prototype requests
Product technologies
Surveys
PRODUCER

# OAIS Model


AIP/SIP templates
AIP/SIP review
Migration packages
Customization advice
Administration
Proposals
Recommendations
Sample records
Develop Preservation Strategies and Standards
Perfomrance info
Consumer comments
Legal and institutional obligations
External data standards
Prototype results
Reports
Define requirements
Alerts
Define basis
Choose records
Identify requirements
Reports
Emerging standards
Approved standards
Mirgration goals
Legal information
Template
Evaluate alternatives
Prototype results
Reports
Go/No-Go
Define alternatives
Alternatives
Develop Packaging Designs & Migration Plans
Develop experiment
Run experiment
Evaluate experiment
Monitor Technology
Prototype results
Preservation requirements
Knowleage Database
Reports
Prototype requests
Prototype requests
Template
Consider results
Analyse results
Set importance factors
Transform measured values
Conclusions
Experience
Reports
Perfomrance info
Consumer comments
Monitor Designated Community
CONSUMER
Surveys
Service requirements
Product technologies
Surveys
PRODUCER

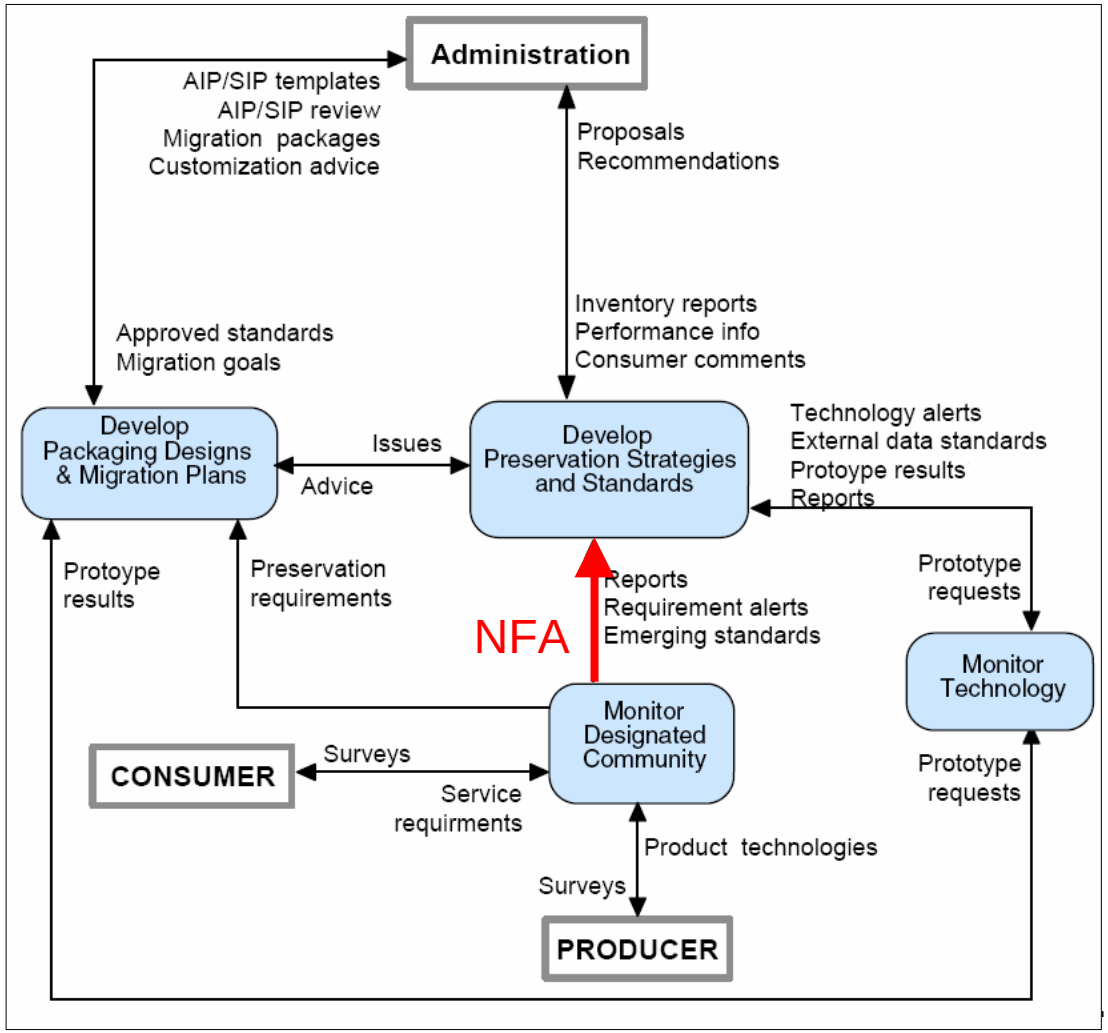
# Alerts

Alerts that can trigger a preservation planning activity

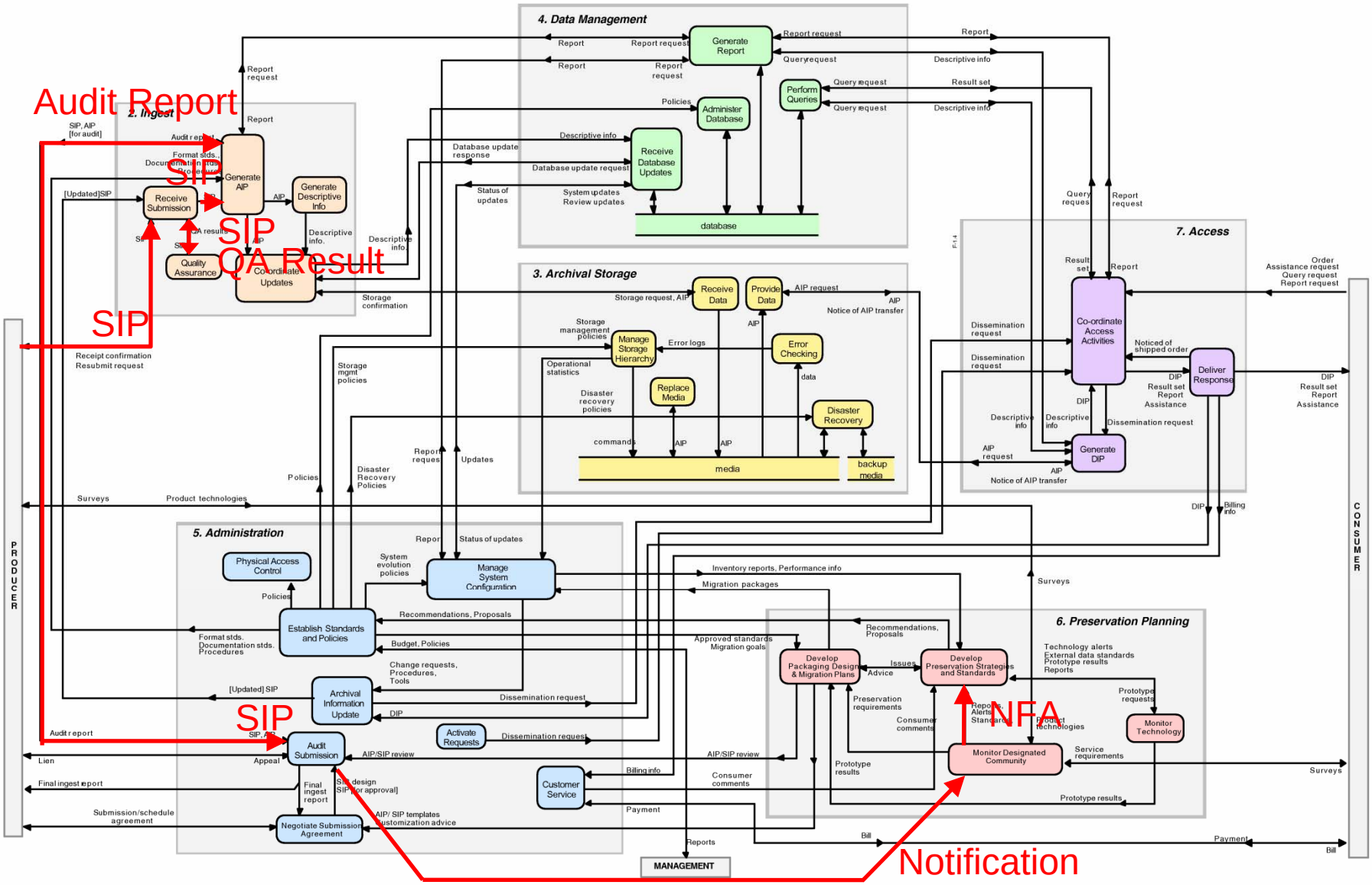
- New Format Alert (NFA)
  - New format is accepted in the archive
  - New preservation strategy has to be identified and evaluated
  - Raised by:
    - Monitor Designated Community
      - Changes in Designated Community
      - Frequent unanticipated submissions

# Change in the Designated Community

For example use of a new file format



# Frequent unanticipated submissions



# Alerts

Alerts that can trigger a preservation planning activity

- New Requirement Alert (NRA)

  – Changes and development in technology
  – Change in preservation requirements
  – Re-evaluation of implemented preservation strategies

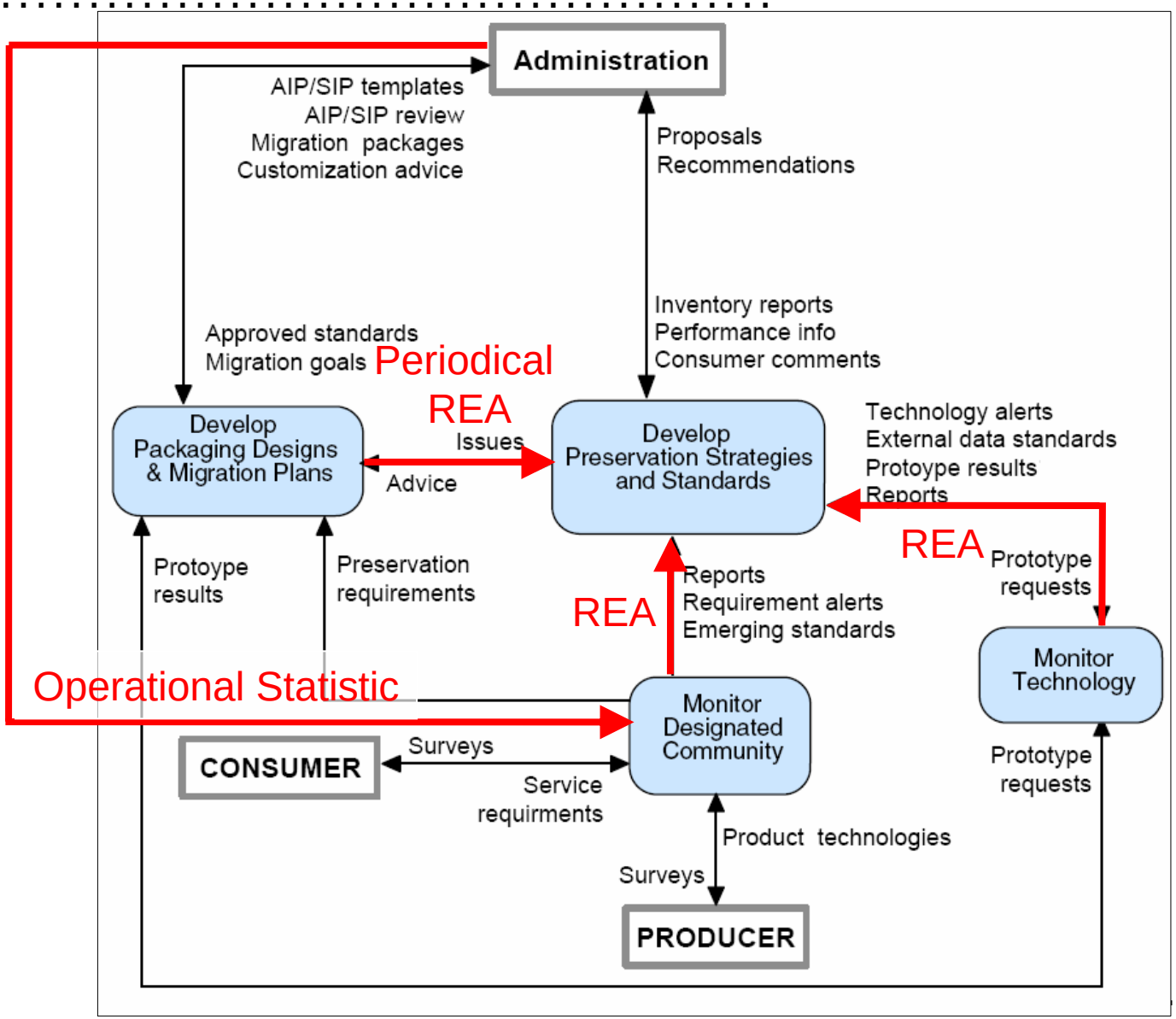
  – Raised by
    - Monitor Designated Community
    - Monitor Technology

# Alerts

Alerts that can trigger a preservation planning activity

- Revision Alert (REA)
  - Periodical reviews help to improve and further develop implemented preservation strategies
  - 3 types:
    - Periodical review
    - Review depending on operational statistics
    - Review for new developments

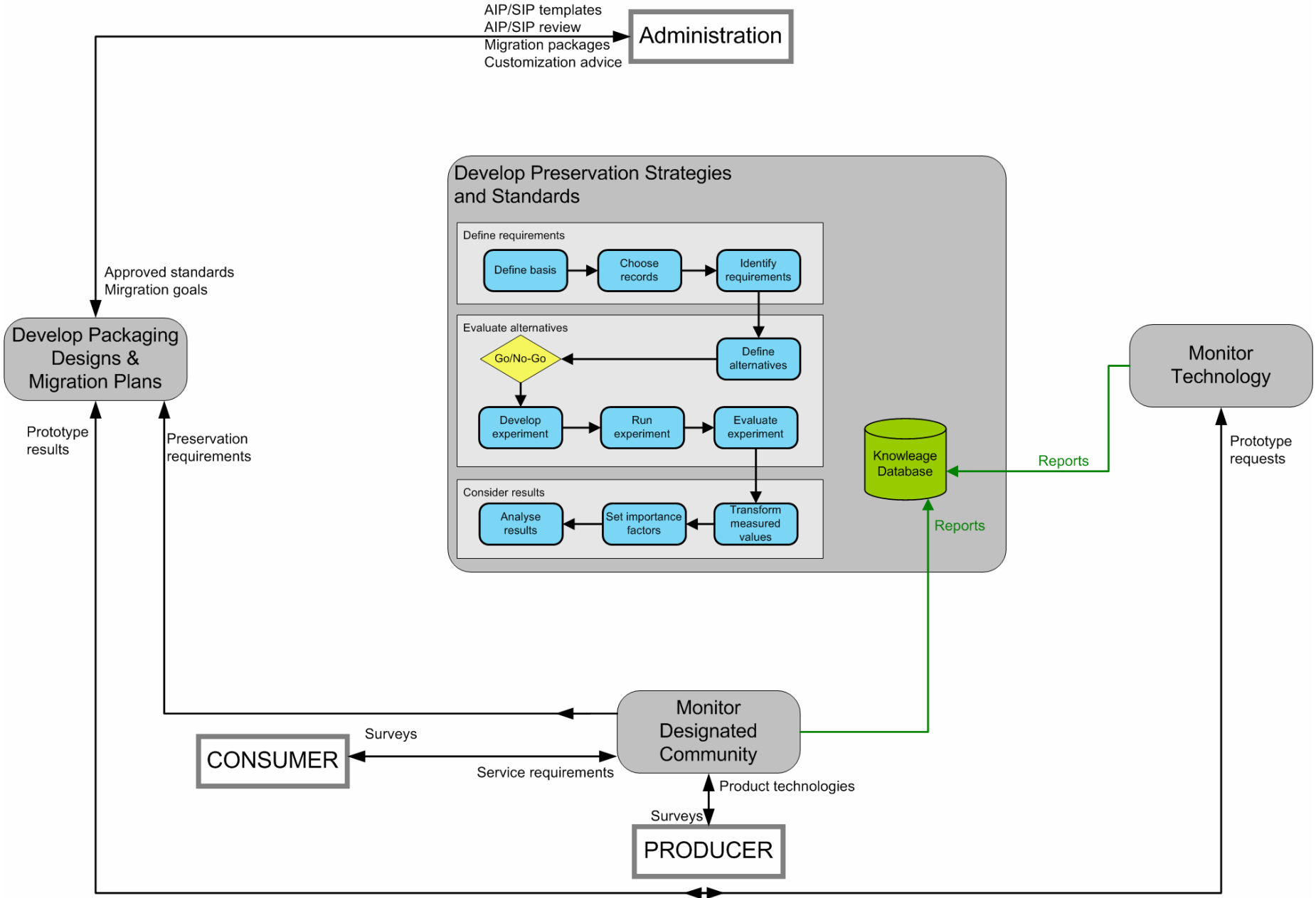
# Revision Alert



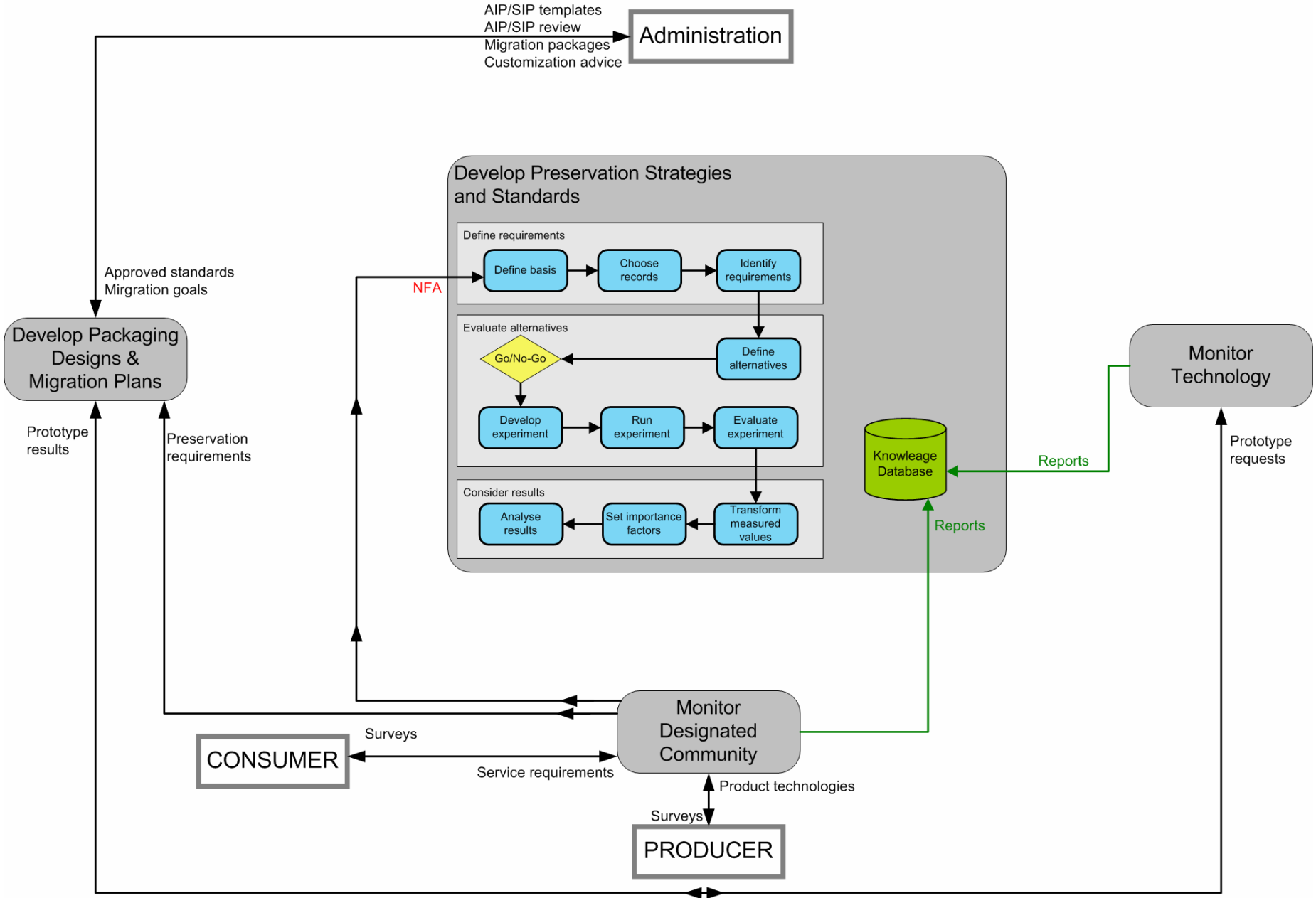
# Preservation Planning Information Flow

- 3 Alerts
  - New File Format Alert
  - New Requirement Alert
  - Revision Alert
- 3 Different Preservation Planning activities
  - Evaluation of new Preservation Strategy for a new format
  - Revision of implemented Preservation Strategies for new requirements
  - Review of implemented Preservation Strategies

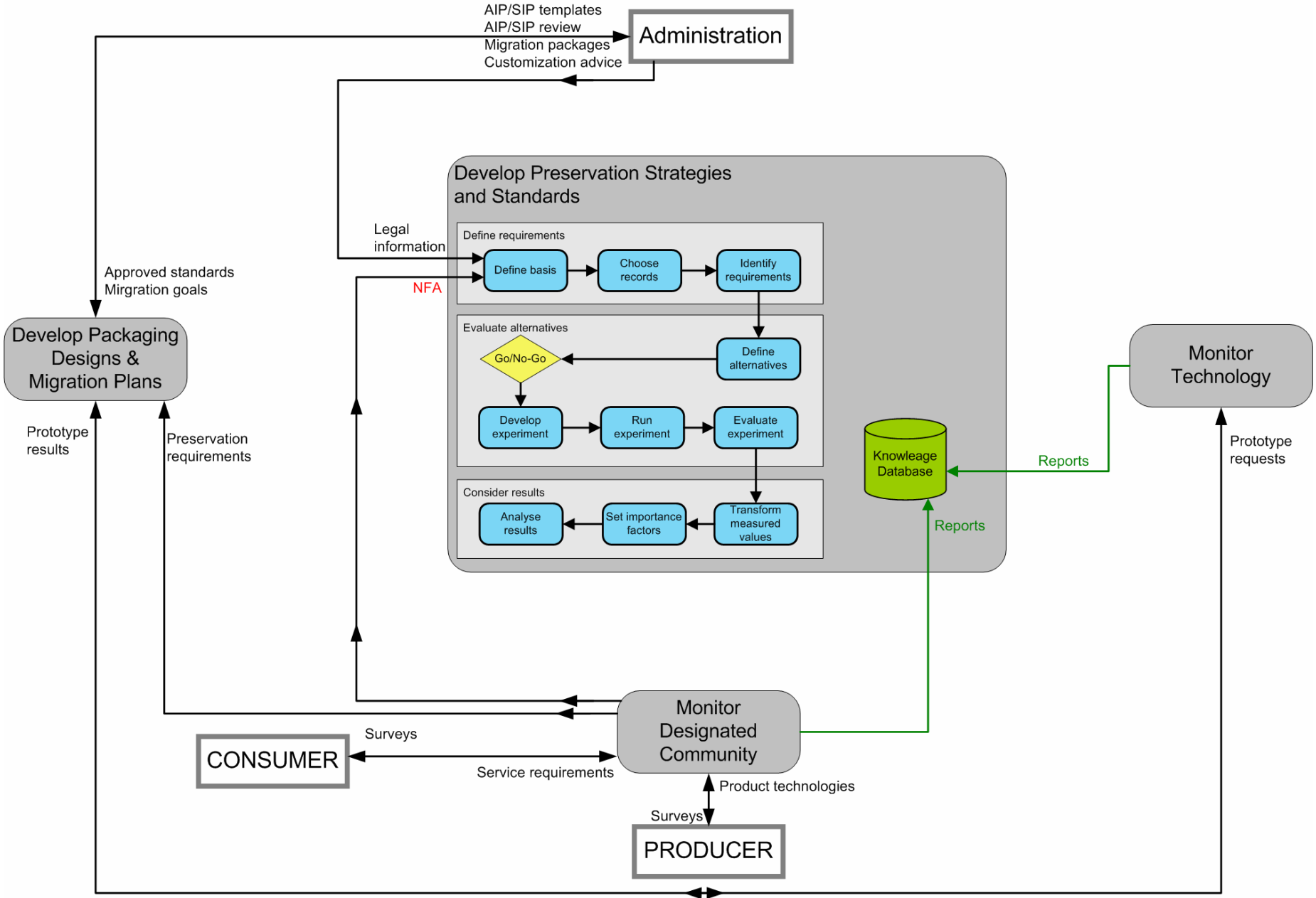
# New Preservation Strategy



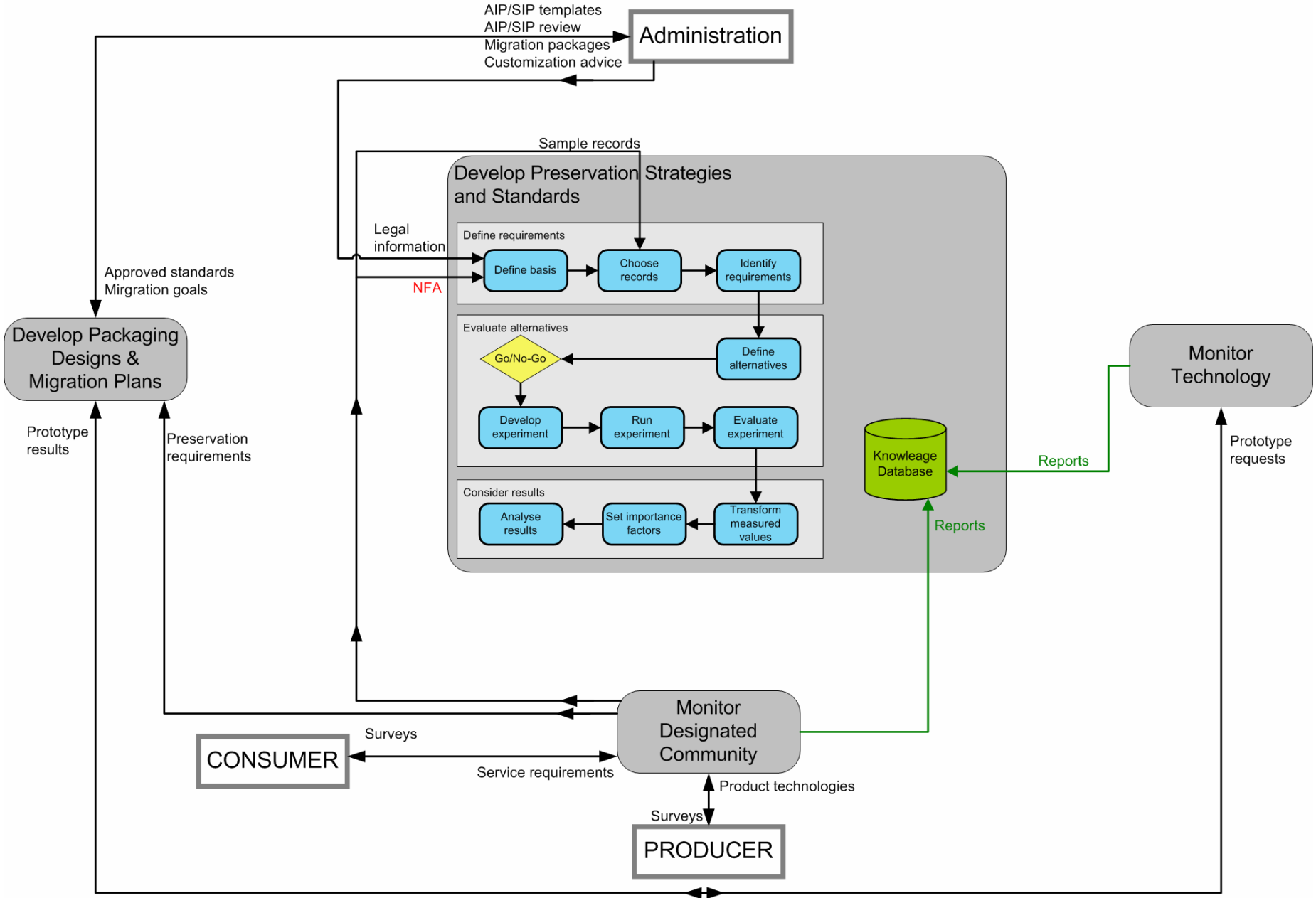
# New Preservation Strategy



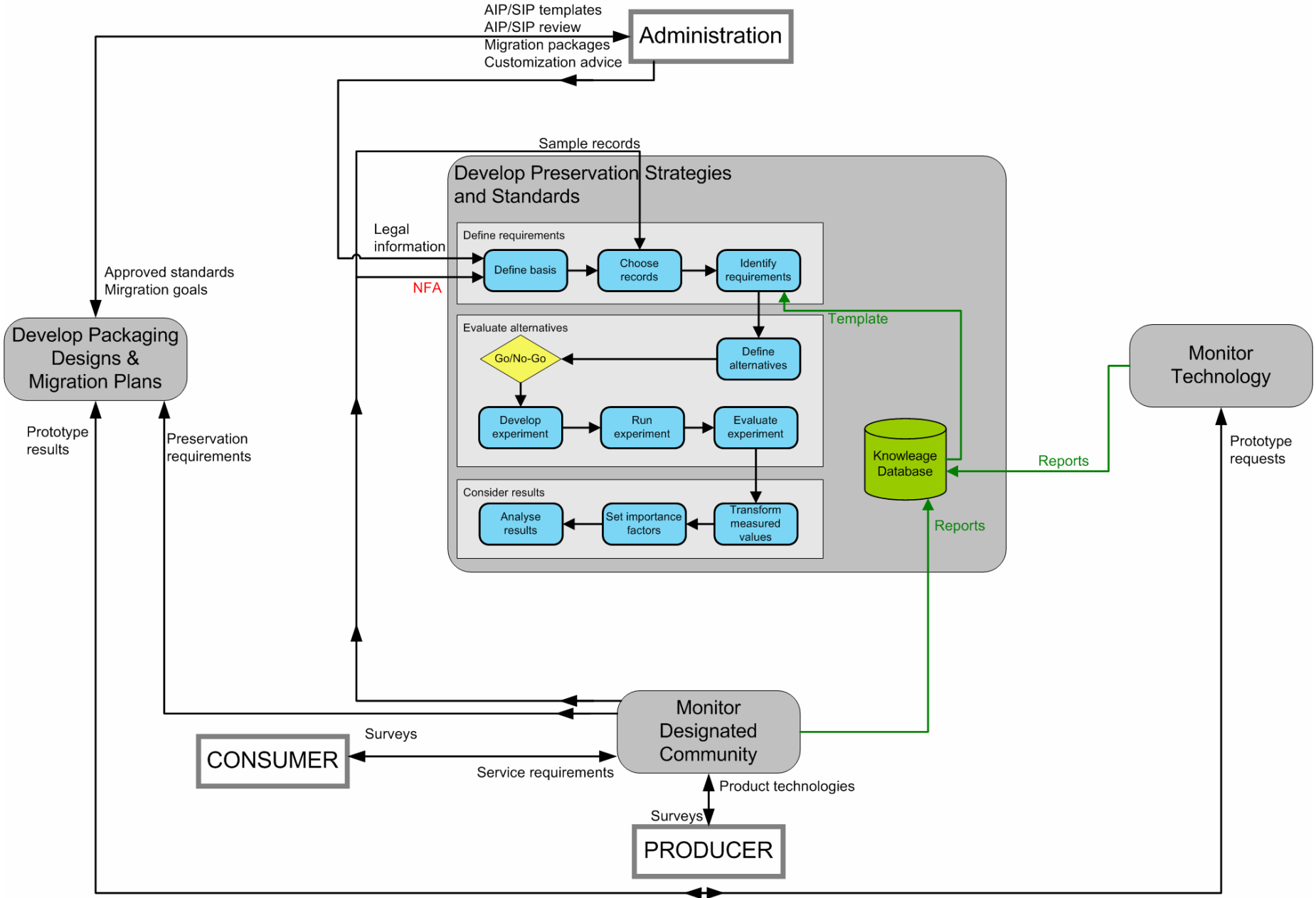
# New Preservation Strategy



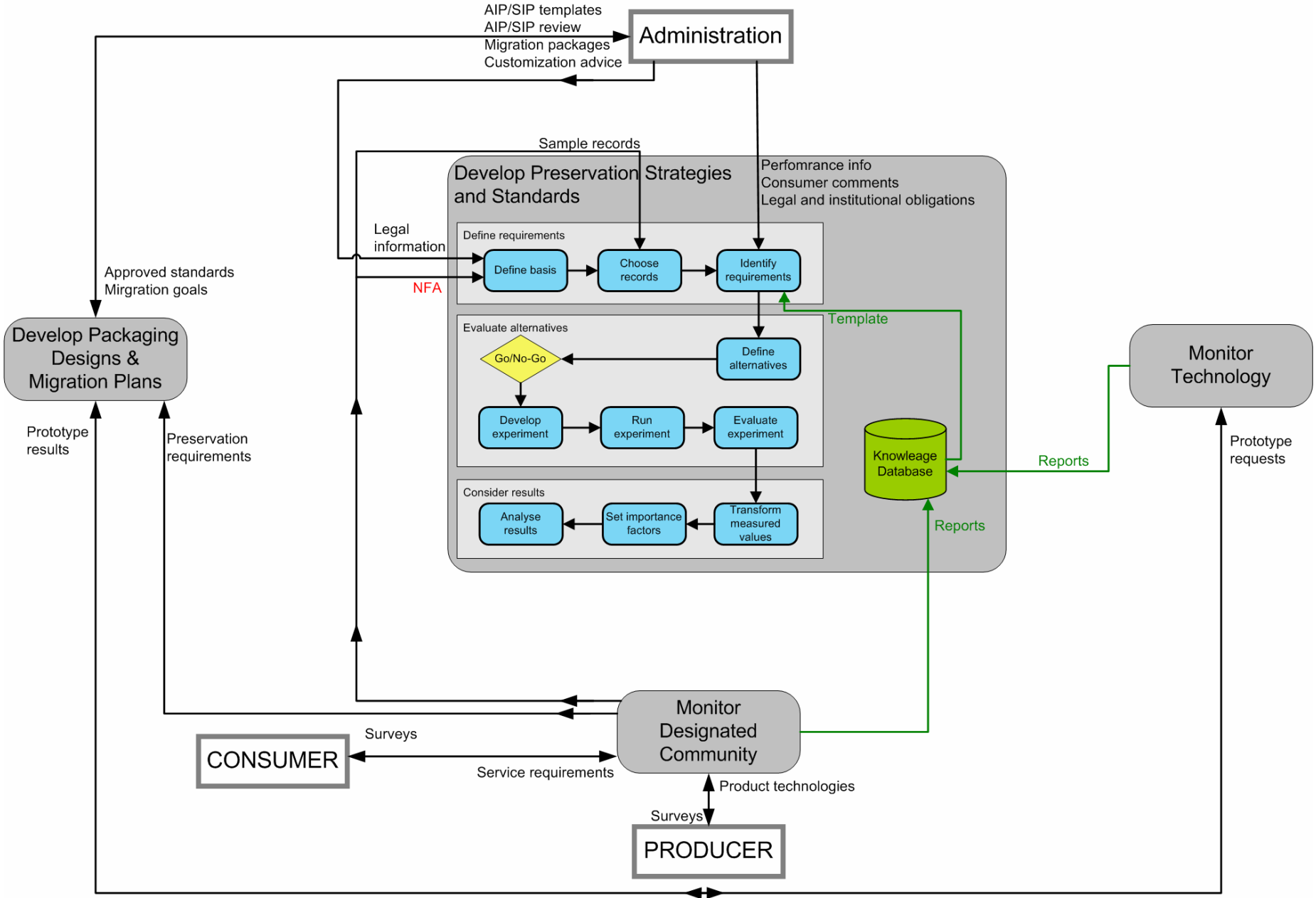
# New Preservation Strategy



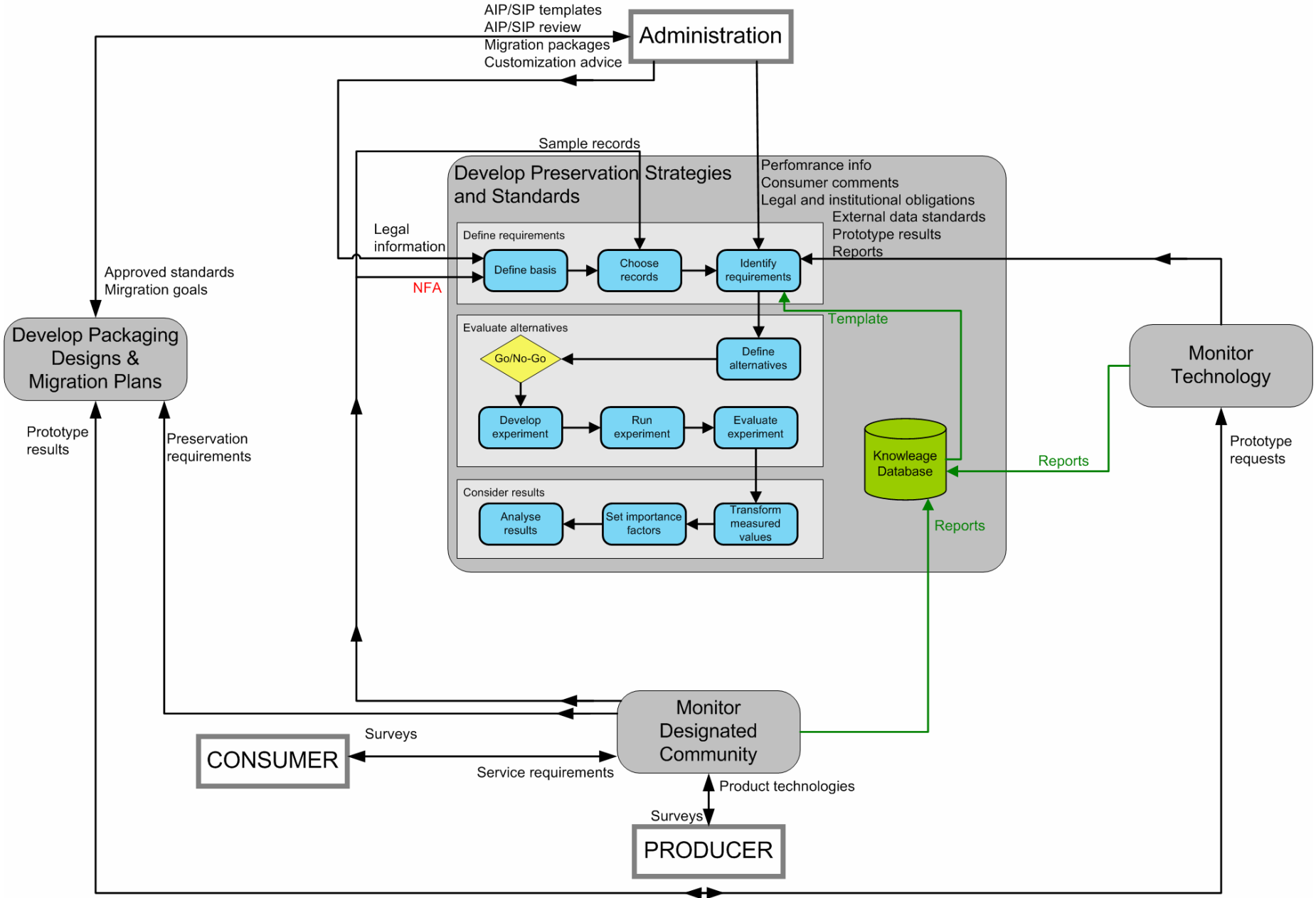
# New Preservation Strategy



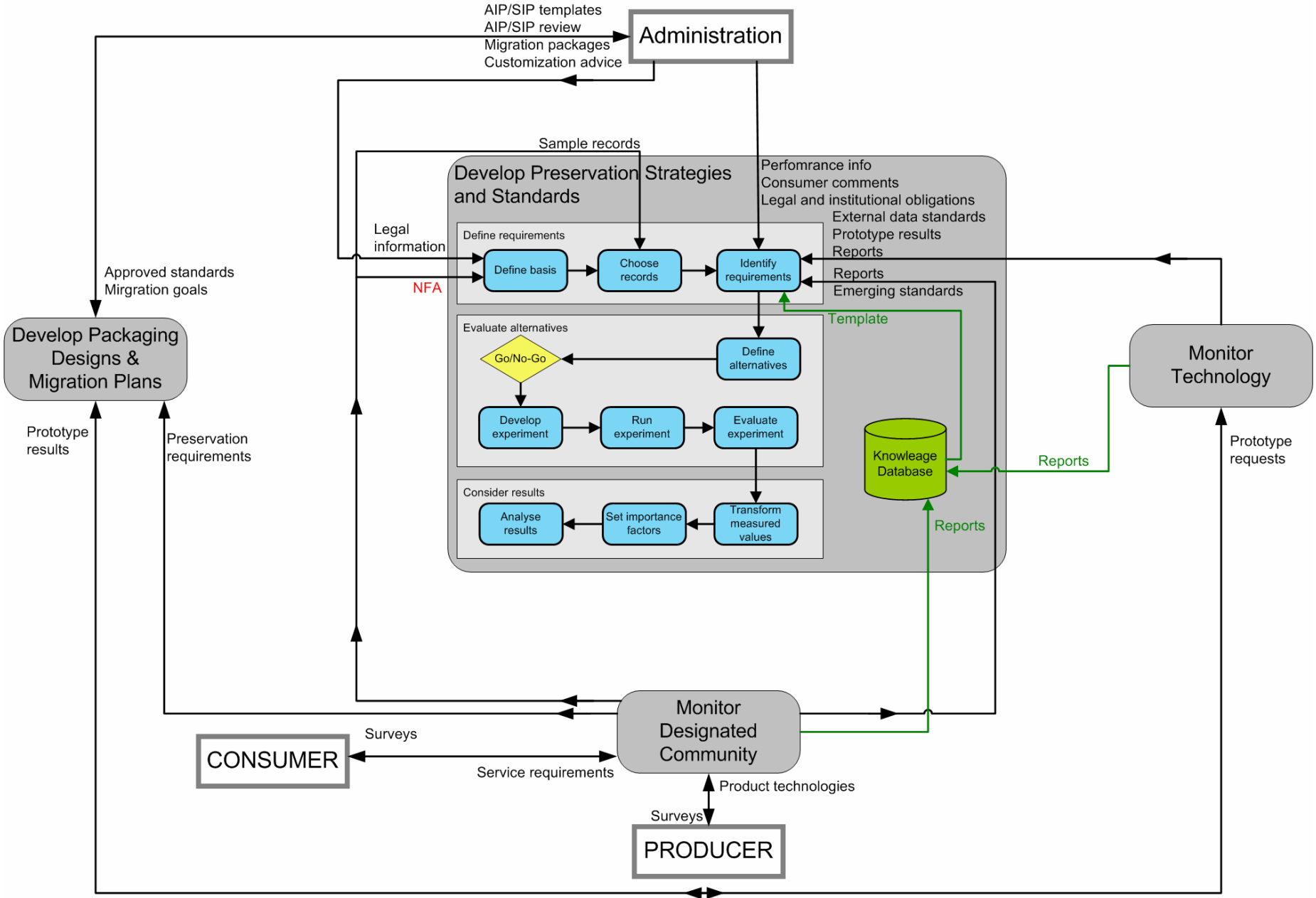
# New Preservation Strategy



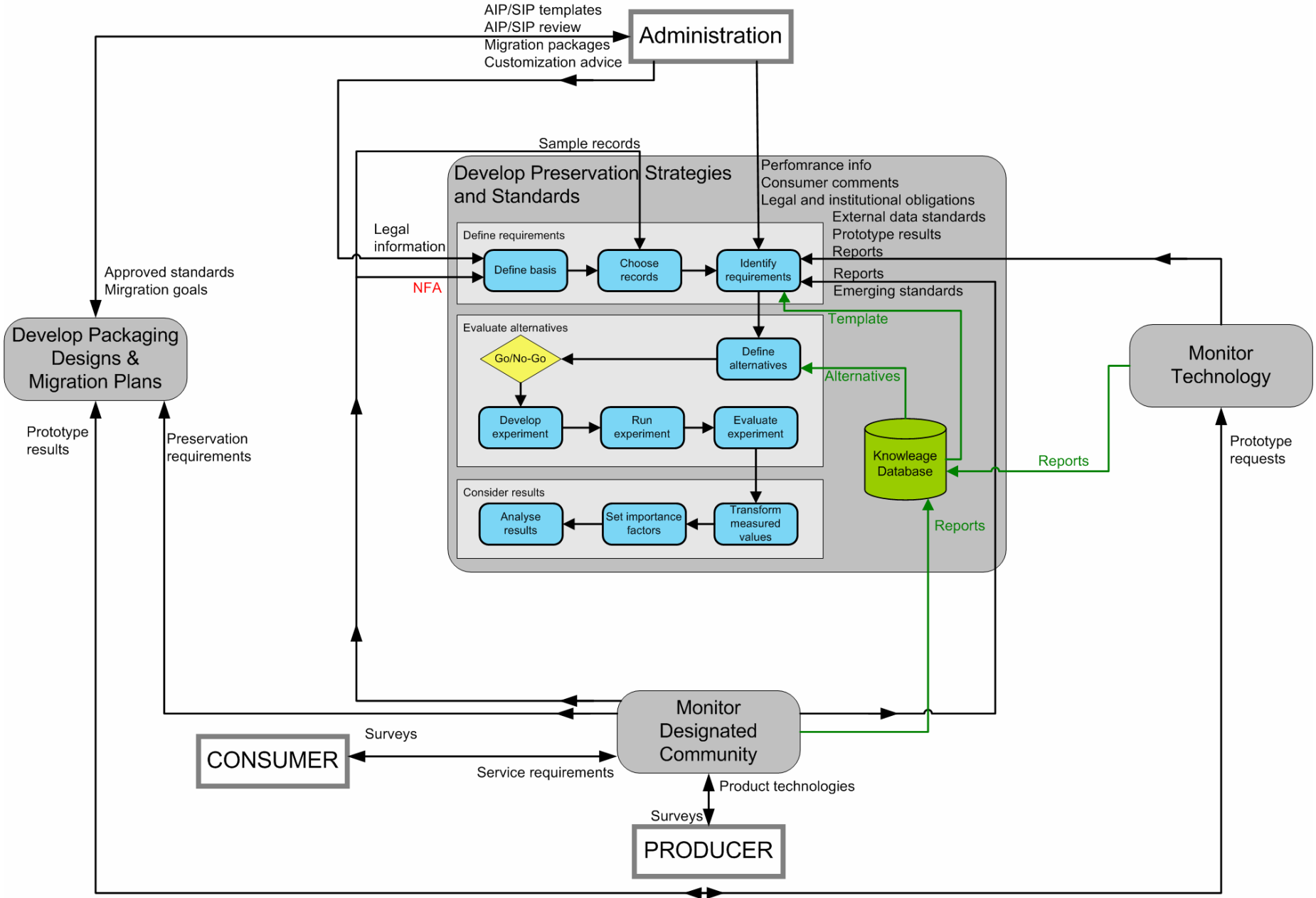
# New Preservation Strategy



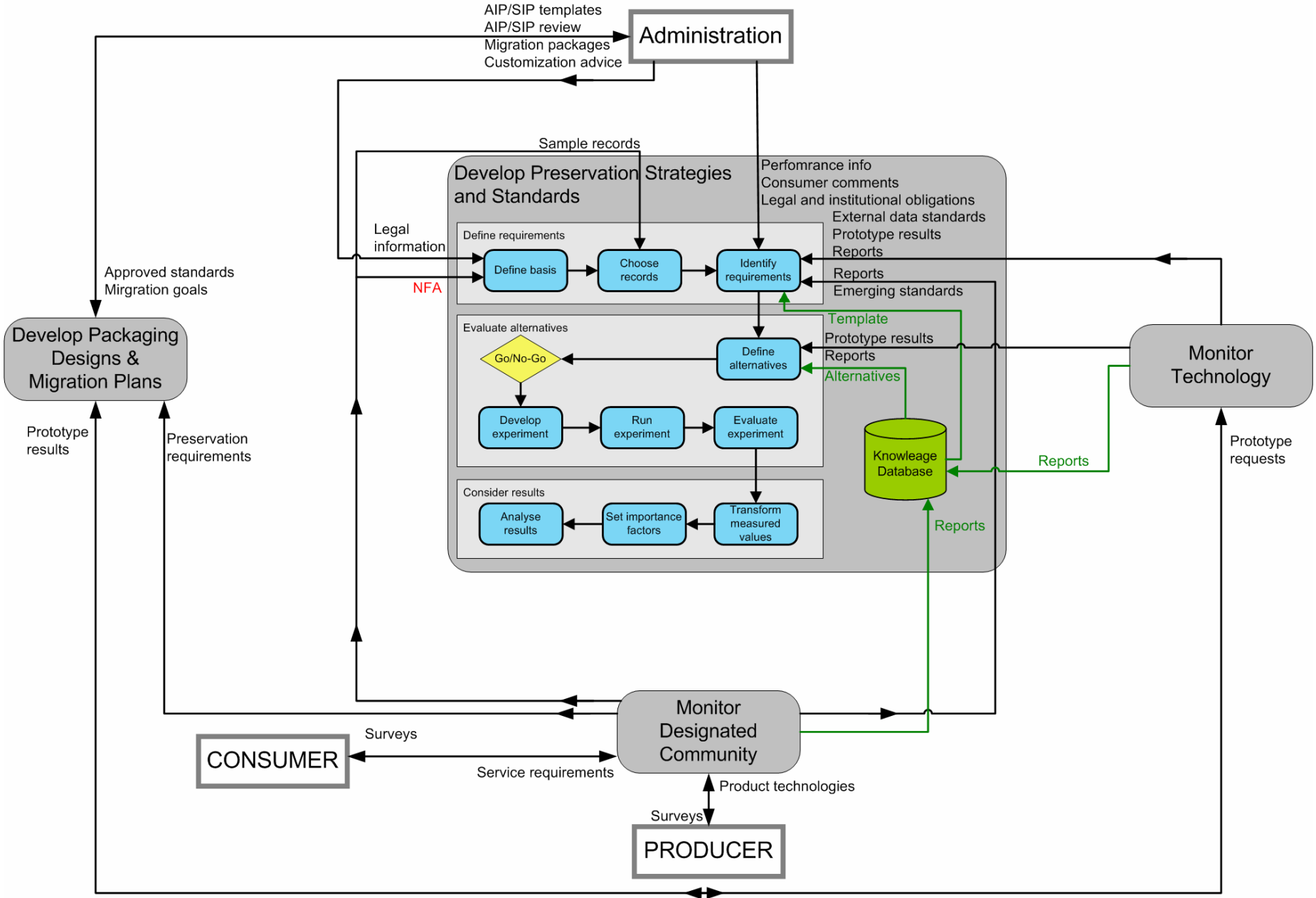
# New Preservation Strategy



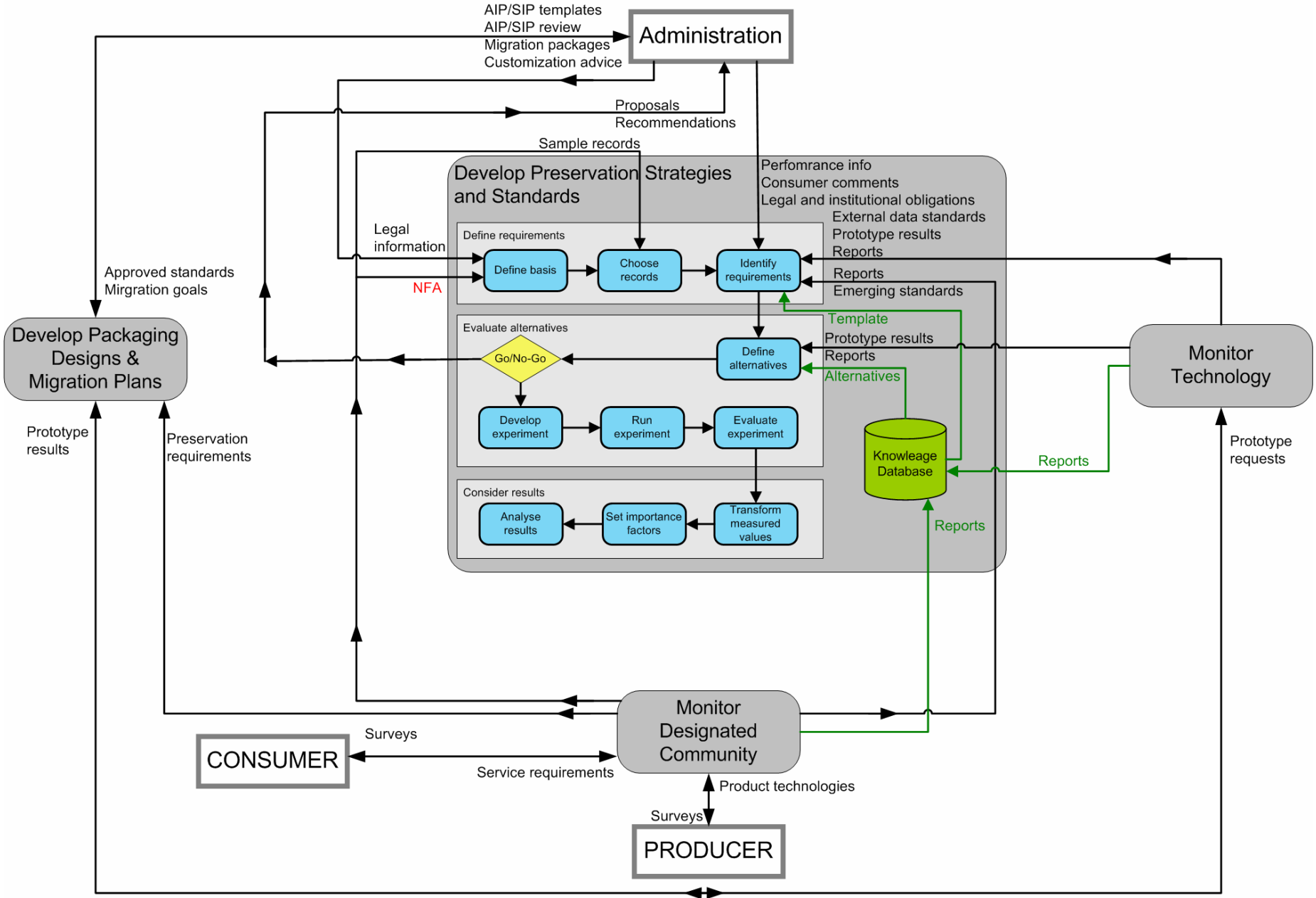
# New Preservation Strategy



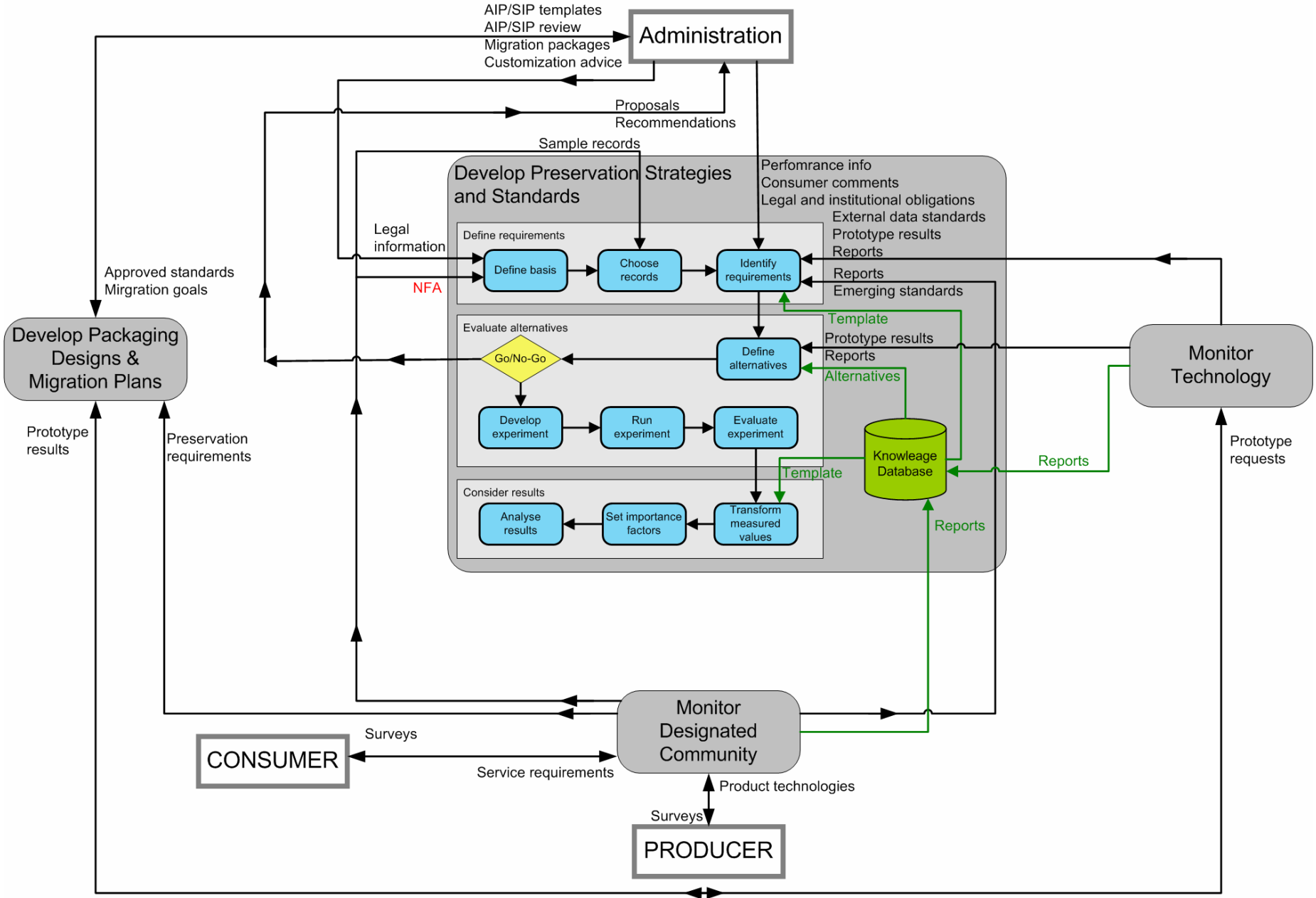
# New Preservation Strategy



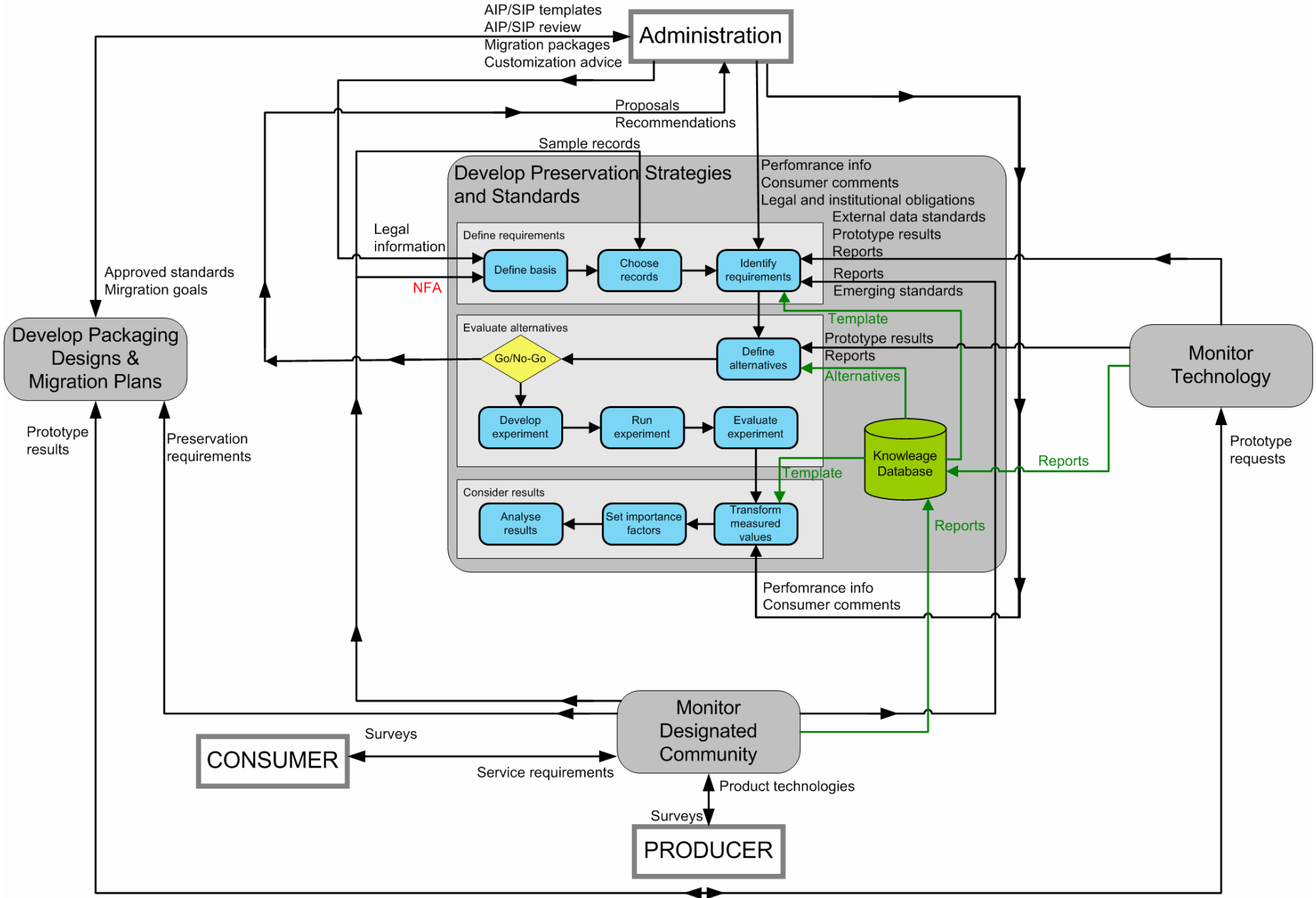
# New Preservation Strategy



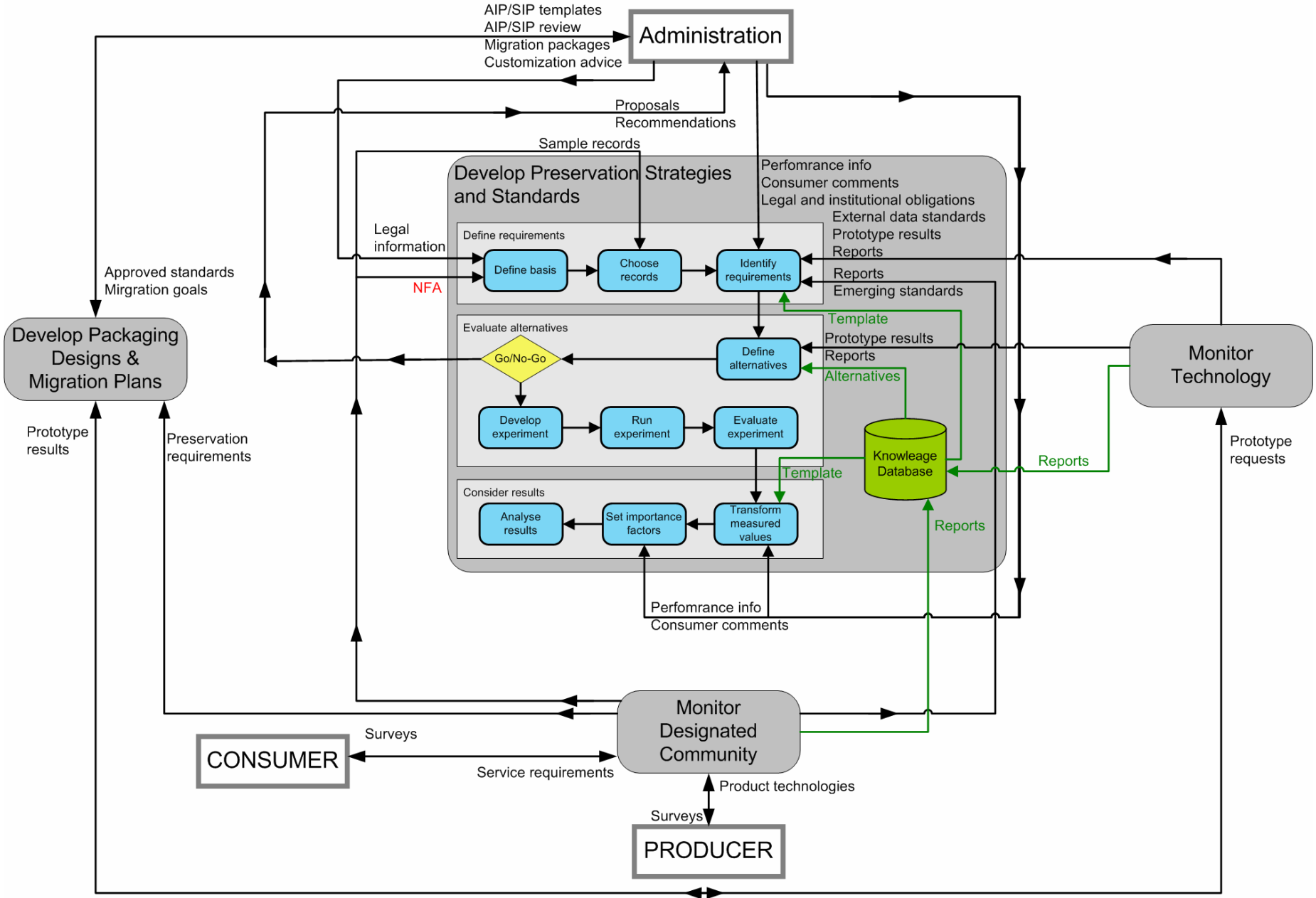
# New Preservation Strategy



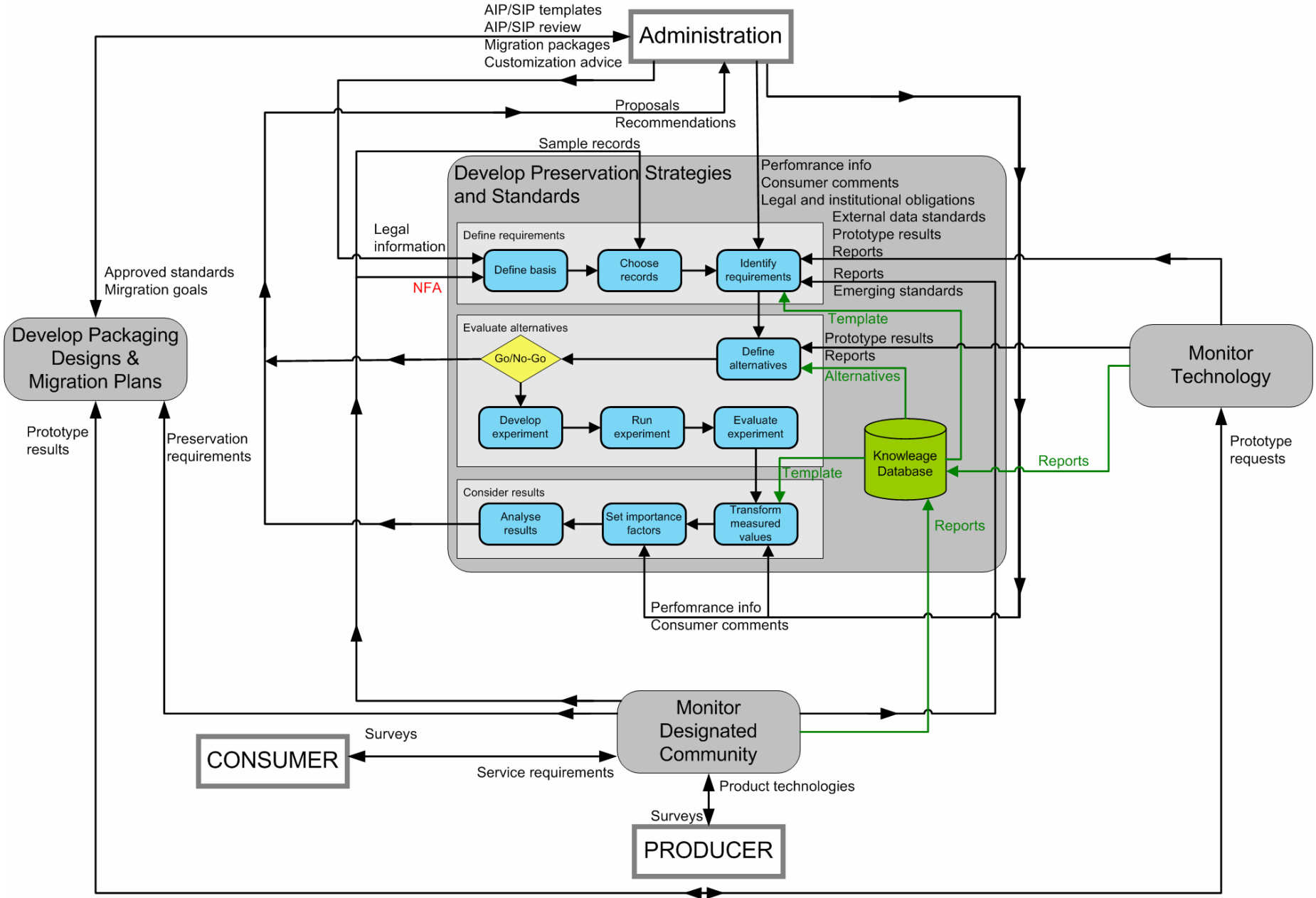
# New Preservation Strategy



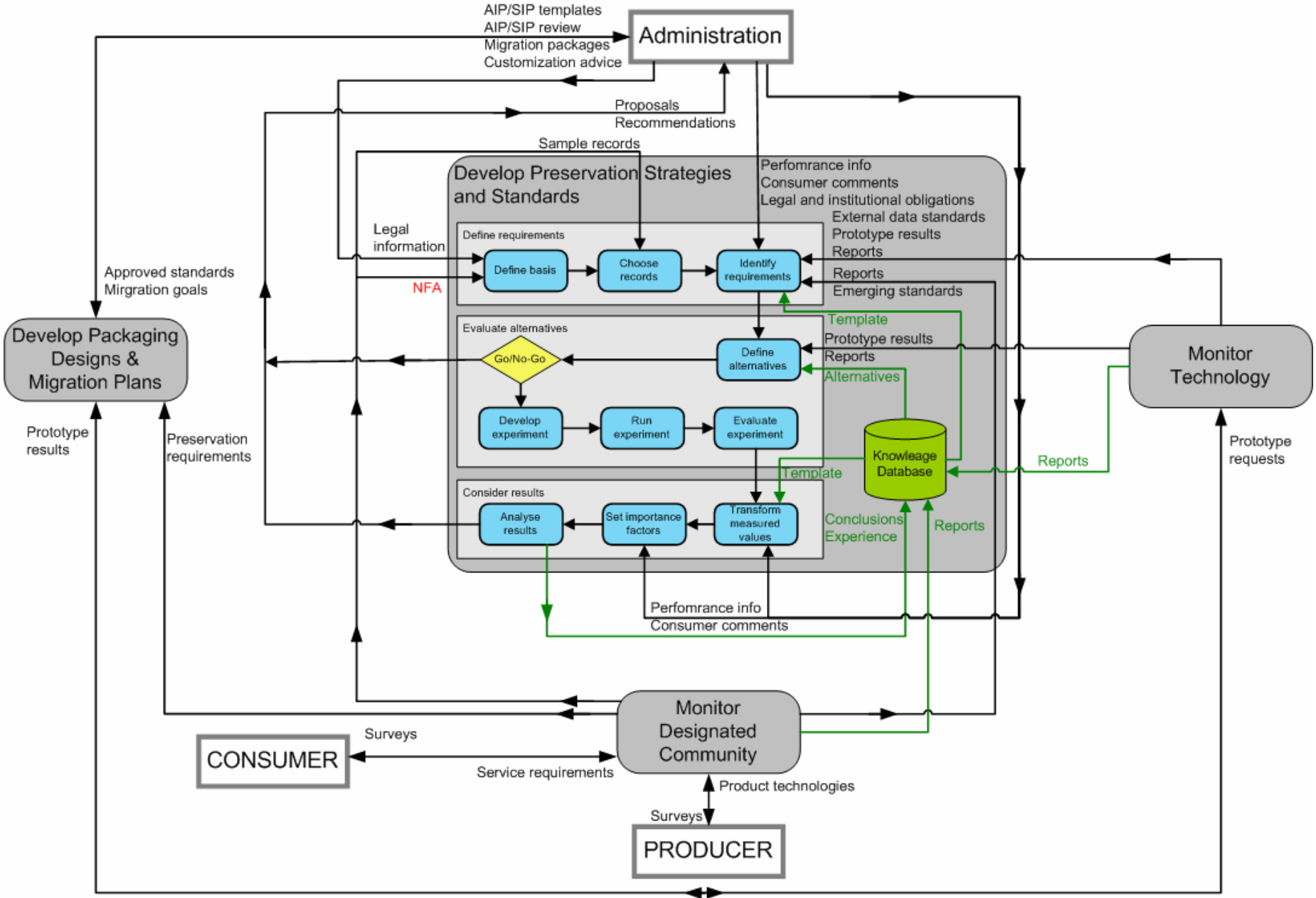
# New Preservation Strategy



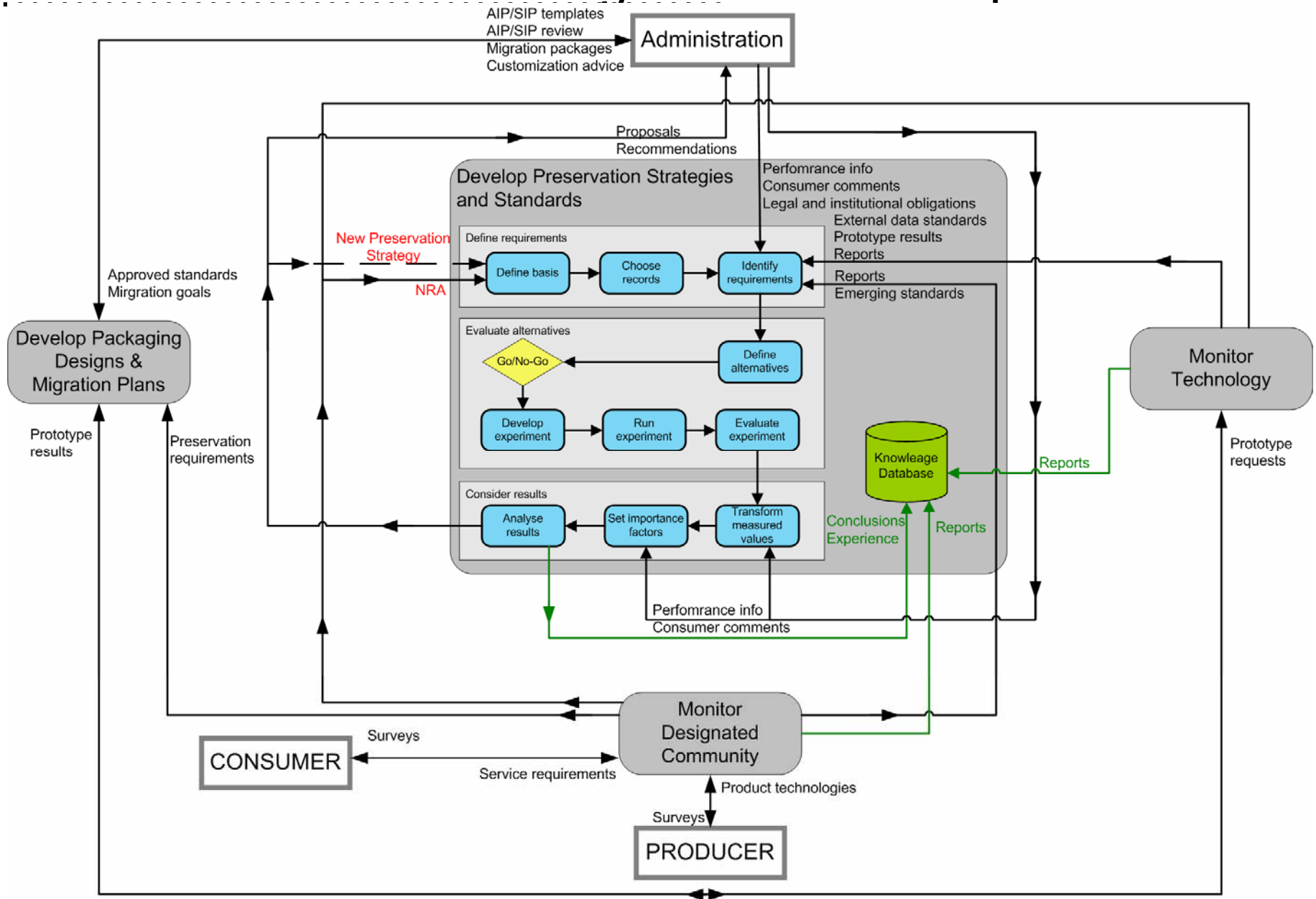
# New Preservation Strategy



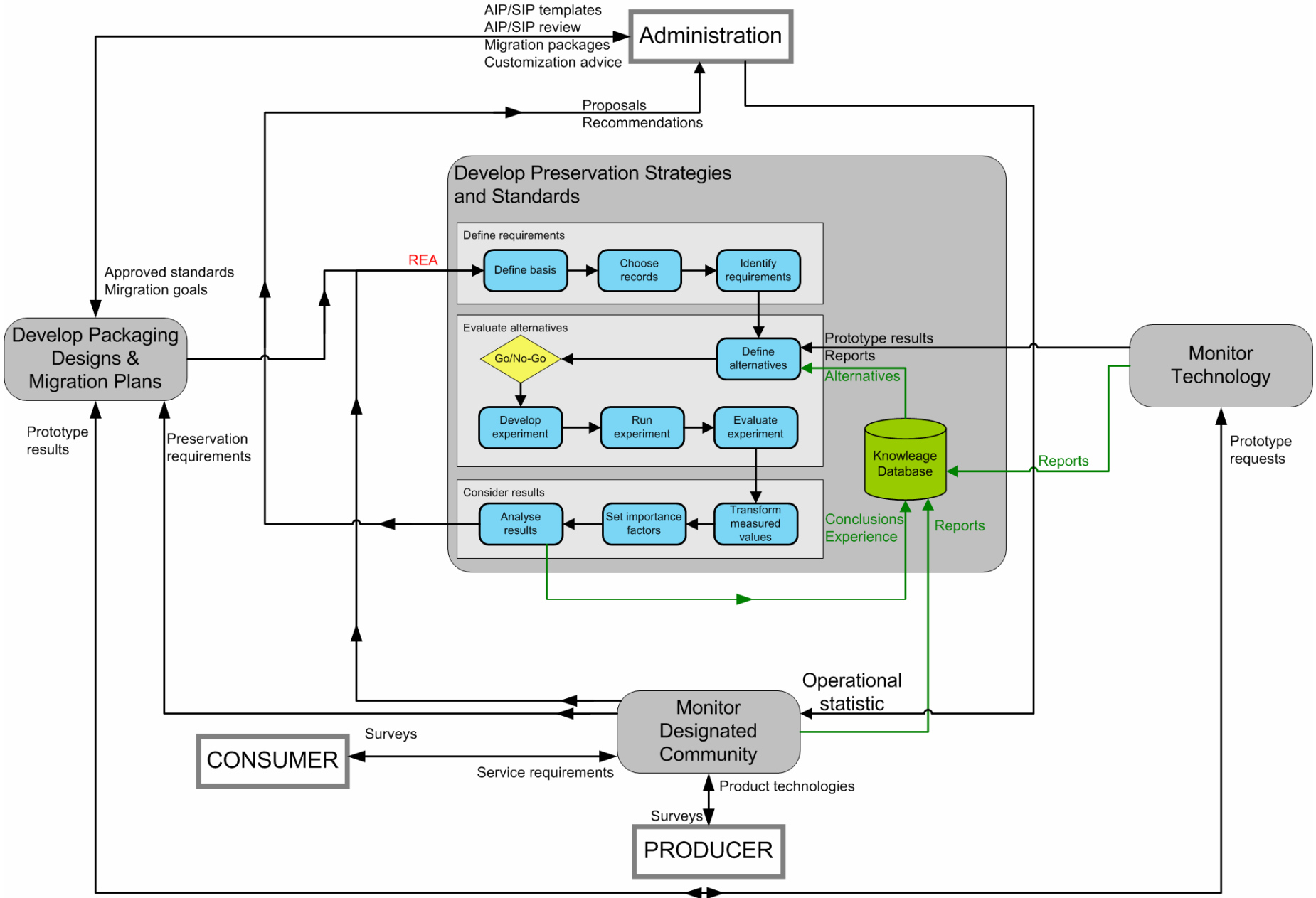
# New Preservation Strategy



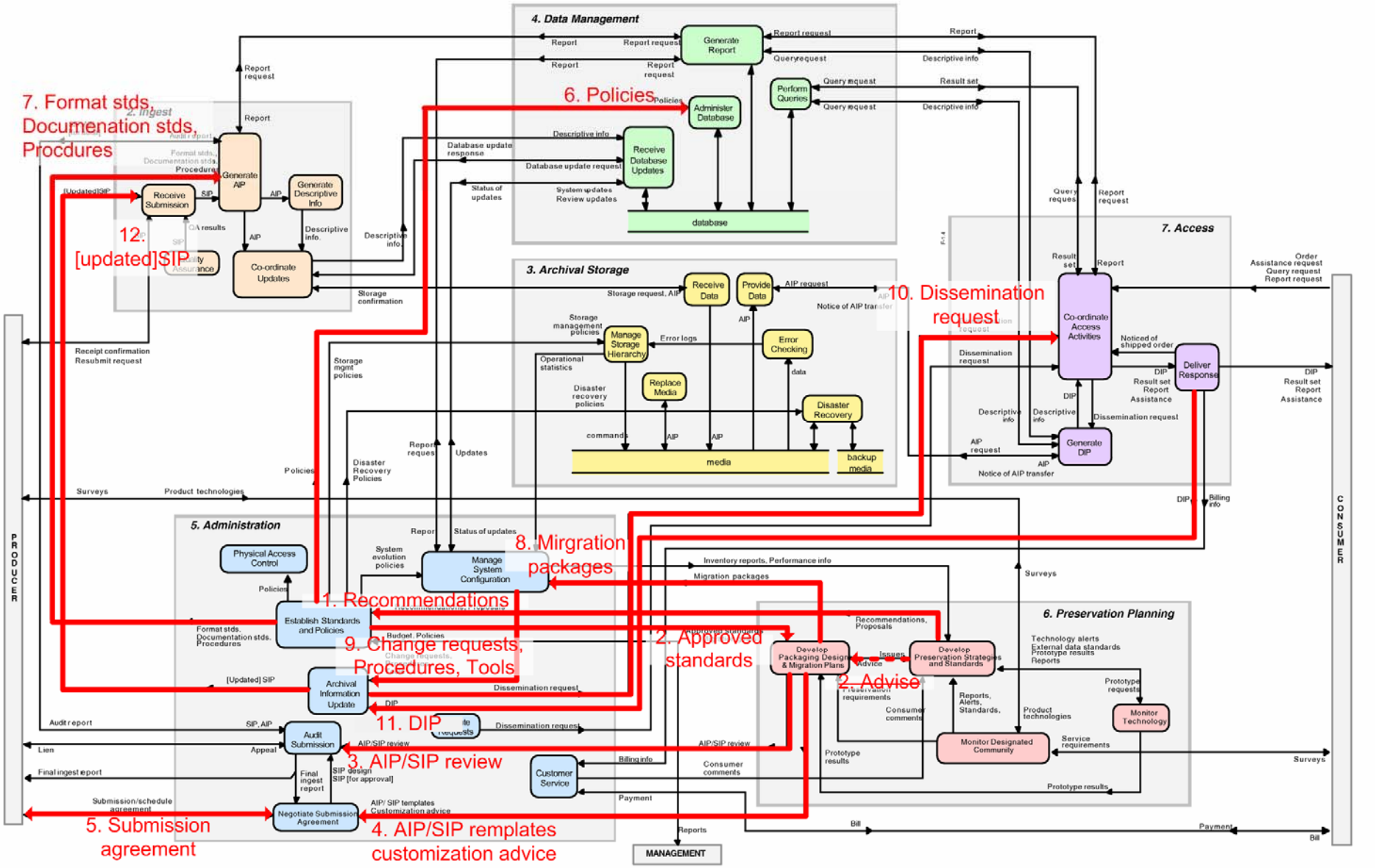
# Revision of existing Preservation Strategies for new Requirements



# Revision Alert



# Effects of New Preservation Strategies



# Conclusion

- More detailed specification of the OAIS Preservation Planning functionality
- We identified
  - 3 Alerts
    - New Format Alert (NFA)
    - New Requirements Alert (NRA)
    - Revision Alert (REA)
  - Information flow for Preservation Planning activities
  - Effects of New Preservation Strategies on other entities

# Tool Support

PLANETS Preservation Planning Tool *(Plato)*

Institute of Software Technology and Interactive Systems

[logout] [help]

Project | Define Requirements ::: Evaluate Requirements ::: Consider Results | Loaded project: PP4 workshop - The National Archive

**Identify Requirements**

Expand All | Collapse All

**Website**

| Focus | Node | Single | Scale | Restriction | Unit |
|---|---|---|---|---|---|
| | ▼ Website | | | | |
| X | ▼ Record characteristics | | | | |
| X | ▶ Appearance | | | | |
| X | ▶ Content | | | | |
| X | ▶ Structure | | | | |
| X | ▼ Behaviour | | | | |
| X | ▼ deactivate | | | | |
| X | ▼ mailto: | ☐ | Ordinal | Yes/No | |
| X | ▼ preserve | | | | |
| X | ▼ menus | ☐ | Ordinal | complete/navigable/missing | |
| X | ▼ pop-ups | ☐ | Ordinal | Yes/No | |
| X | ▼ freeze | | | | |
| X | ▼ current date/time | ☐ | Ordinal | frozen/missing/current | |
| X | ▼ visitor counter | ☐ | Ordinal | frozen/missing/current | |
| X | ▼ Newsfeeds | ☐ | Ordinal | frozen/missing/current | |
| X | ▶ Context | | | | |
| X | ▼ Technical characteristics | | | | |
| X | ▼ Ubiquity | ☐ | Ordinal | Ubiquitous/Widespread/Specialised/Obs | |
| X | ▼ Tool Support | ☐ | Positive Number | | Number of tools |
| X | ▼ Documentation | | | | |
| X | ▼ Quality | ☐ | Ordinal | Primary/Secondary | |
| X | ▼ Disclosure | ☐ | Ordinal | Full/Partial/None | |
| X | ▼ Openness | ☐ | Ordinal | Standard/Open/Proprietary | |
| | | | Ordinal | Public/Limited/None | |

© 2007 Institute of Software Technology and Interactive Systems: «office bears»

Quick Access:

# PLANETS

- Four year Digital Preservation Project funded by the European Commission
- 16 Partners from Libraries and Archives, research institutions and technology companies
- Objectives
  - Preservation Action
  - Preservation Characterisation
  - Preservation Planning
  - Interoperability Framework
  - Testbed
- More information about Planets:

  www.planets-project.eu