

De Cabinda ao Cunene...

Preservação Digital - da teoria à prática

Ciclo de reflexões do Grupo de Trabalho de Gestão de Documentos de Arquivo da BAD

12 Julho 2013 // Arquivo Distrital do Porto

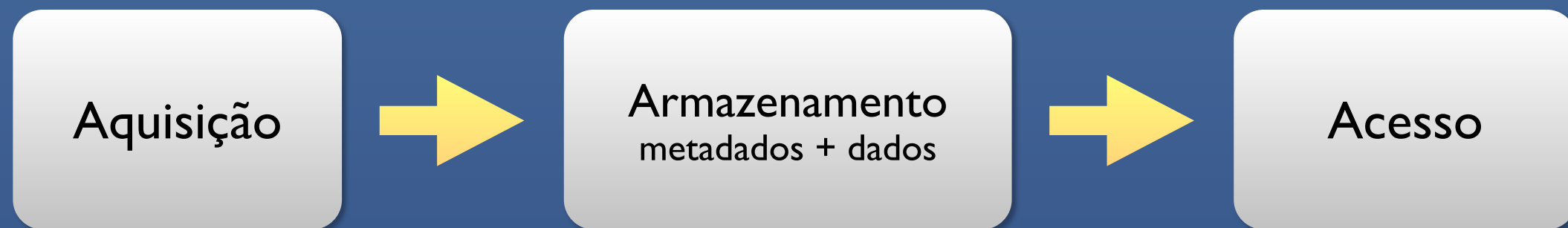
Miguel Ferreira, PhD.

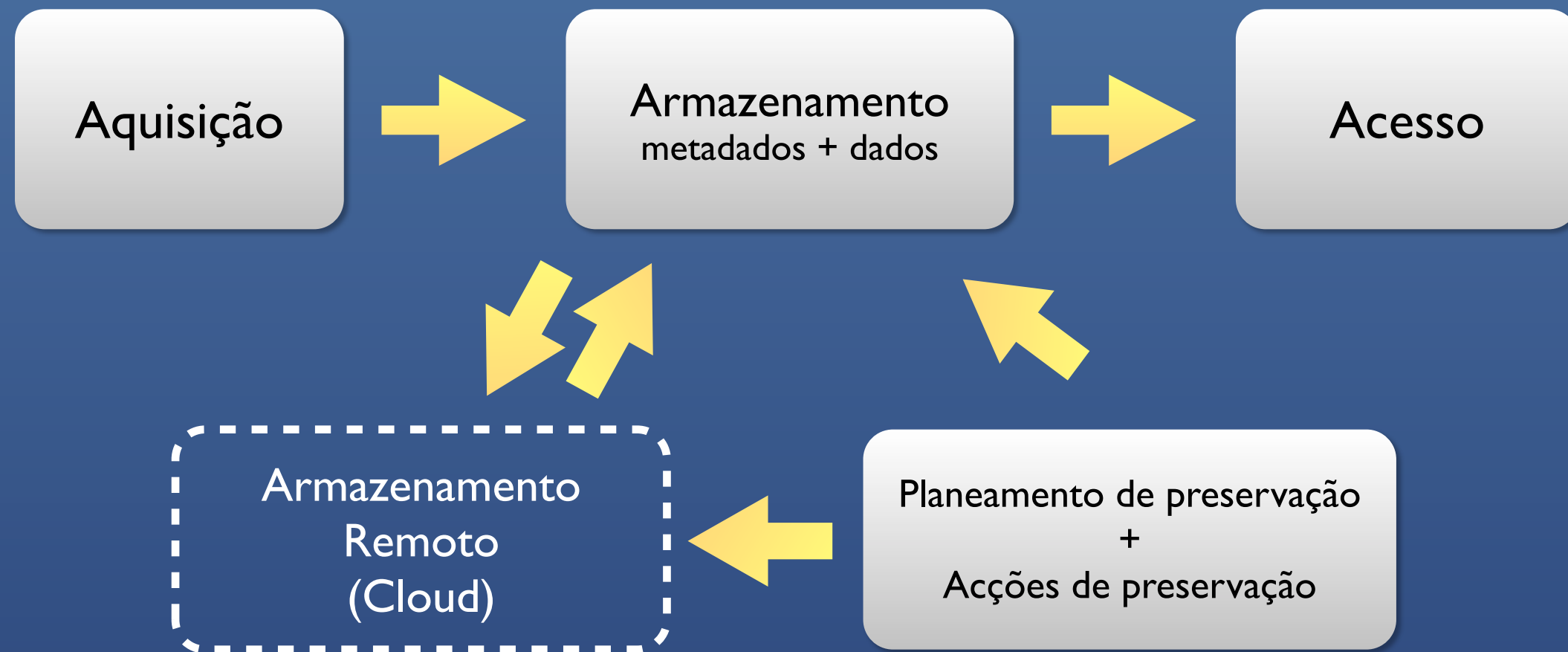
Director executivo // KEEP SOLUTIONS

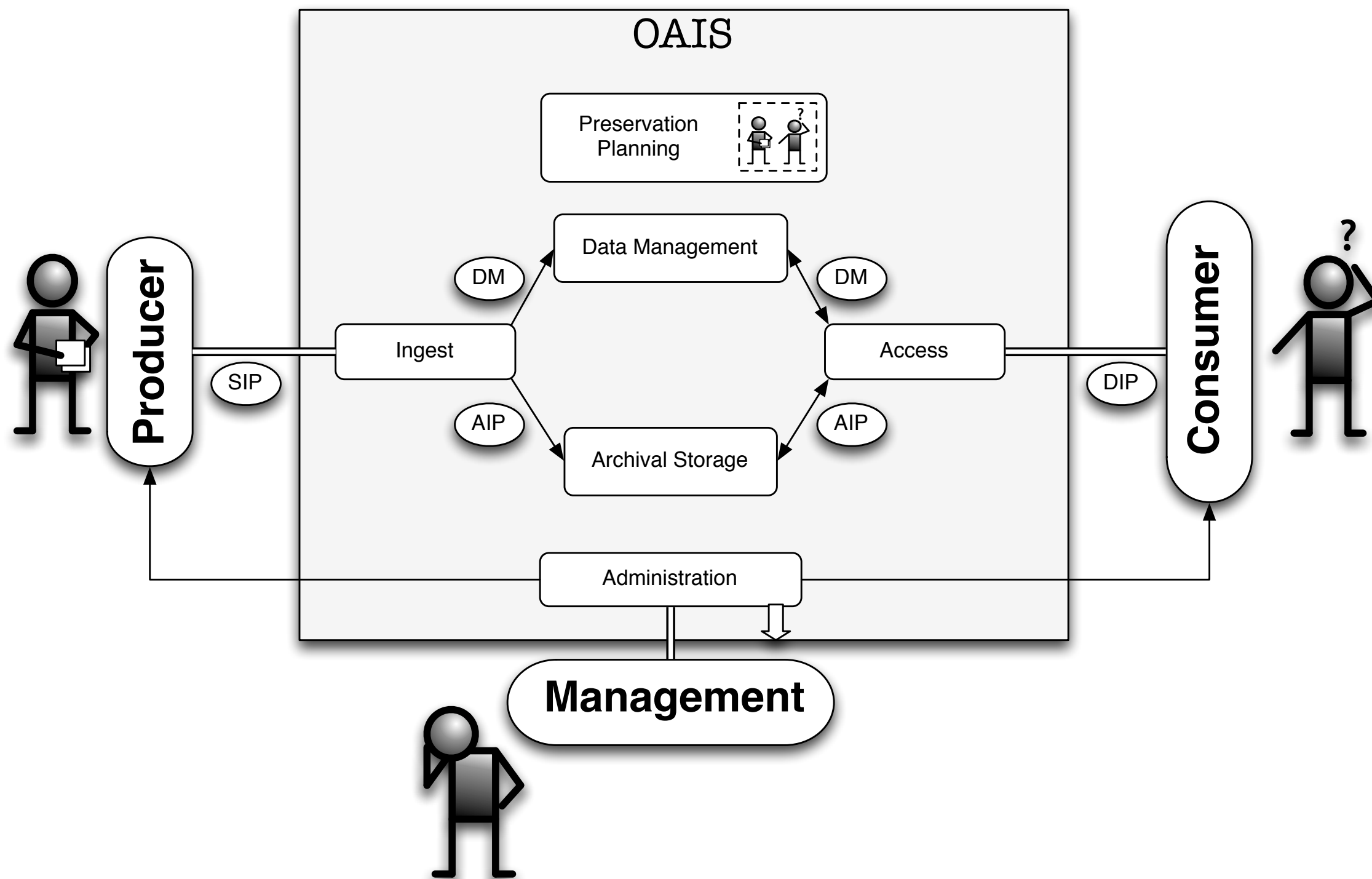
mferreira@keep.pt

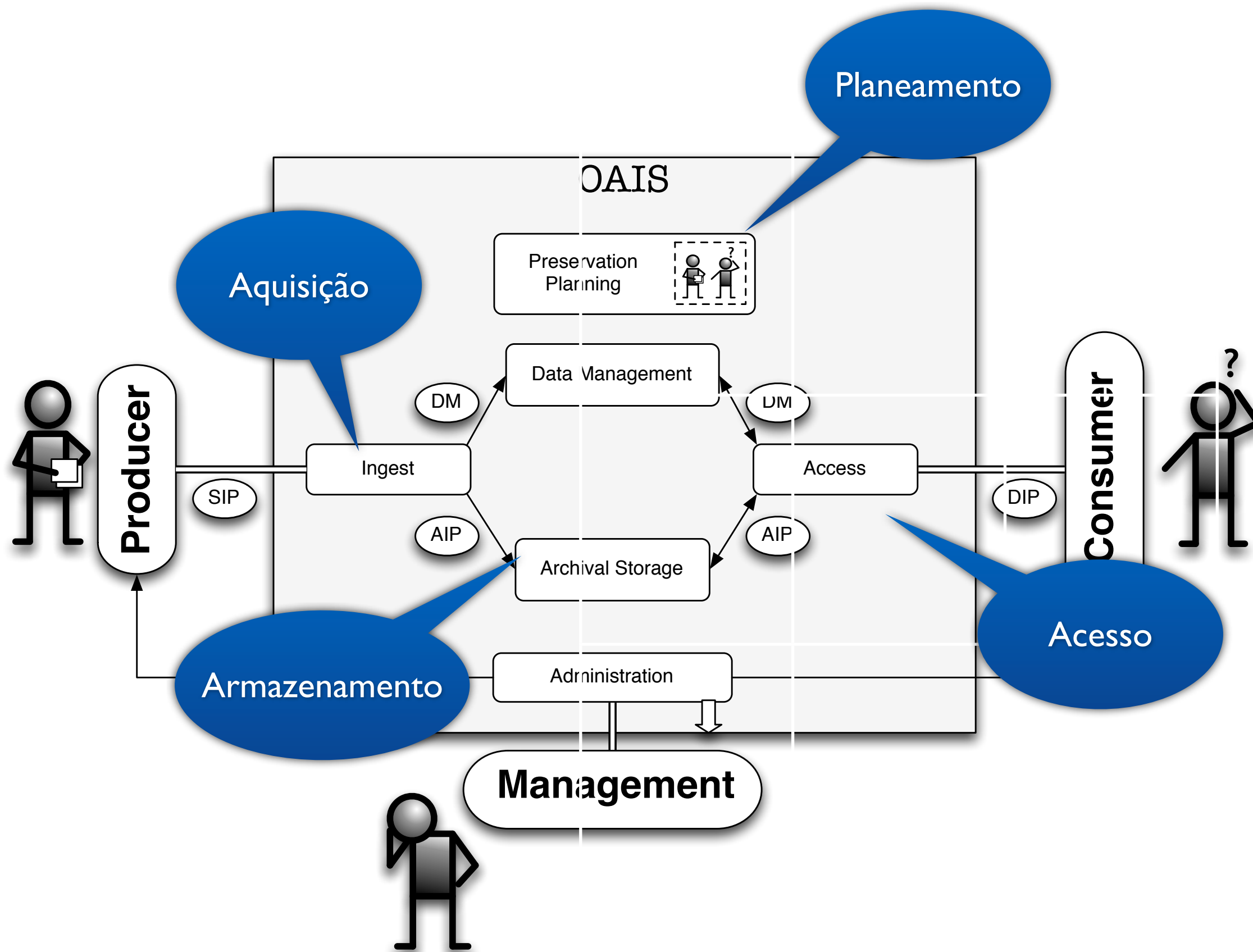
Do que vamos falar?

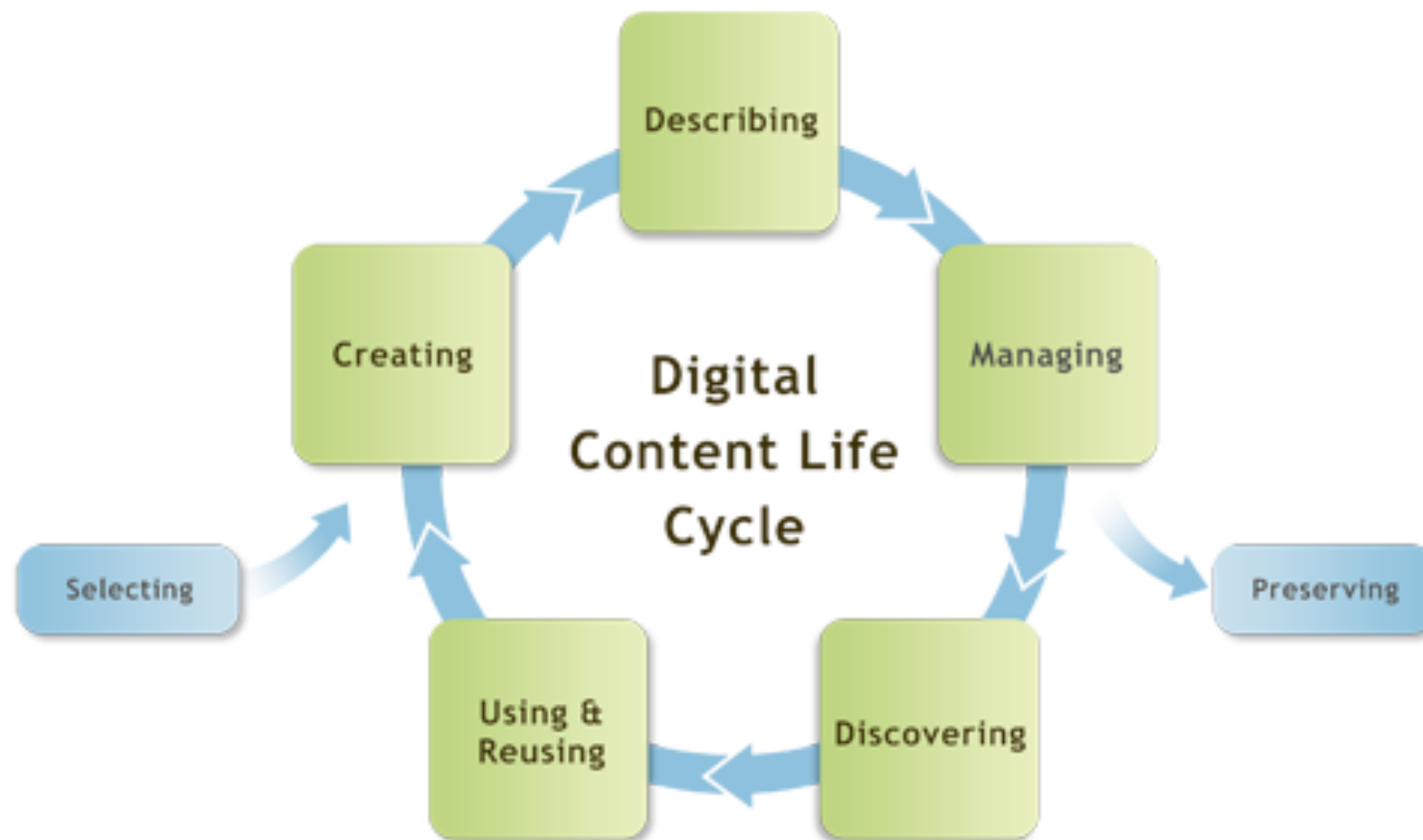
Ciclo de vida de um objecto digital



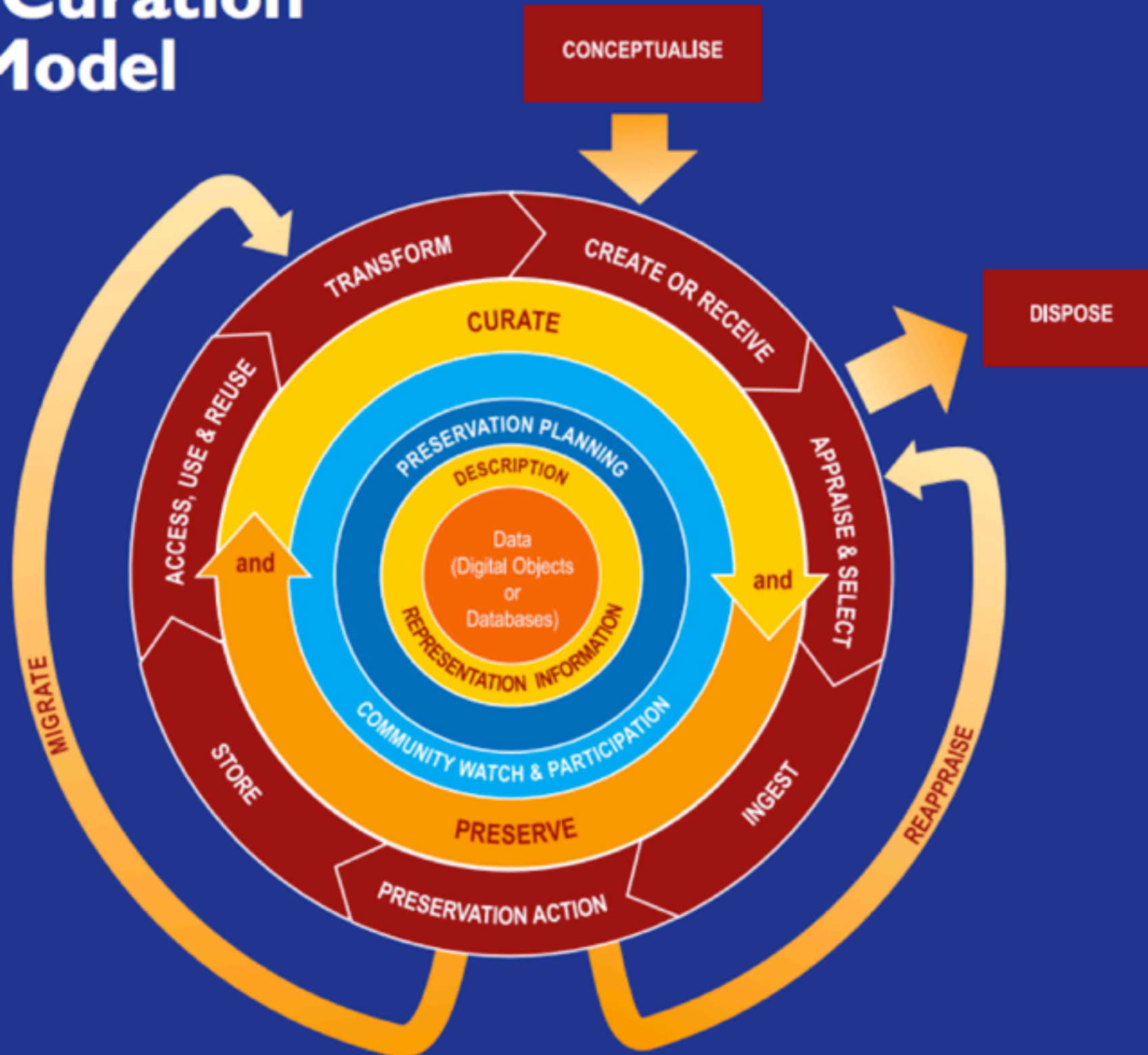




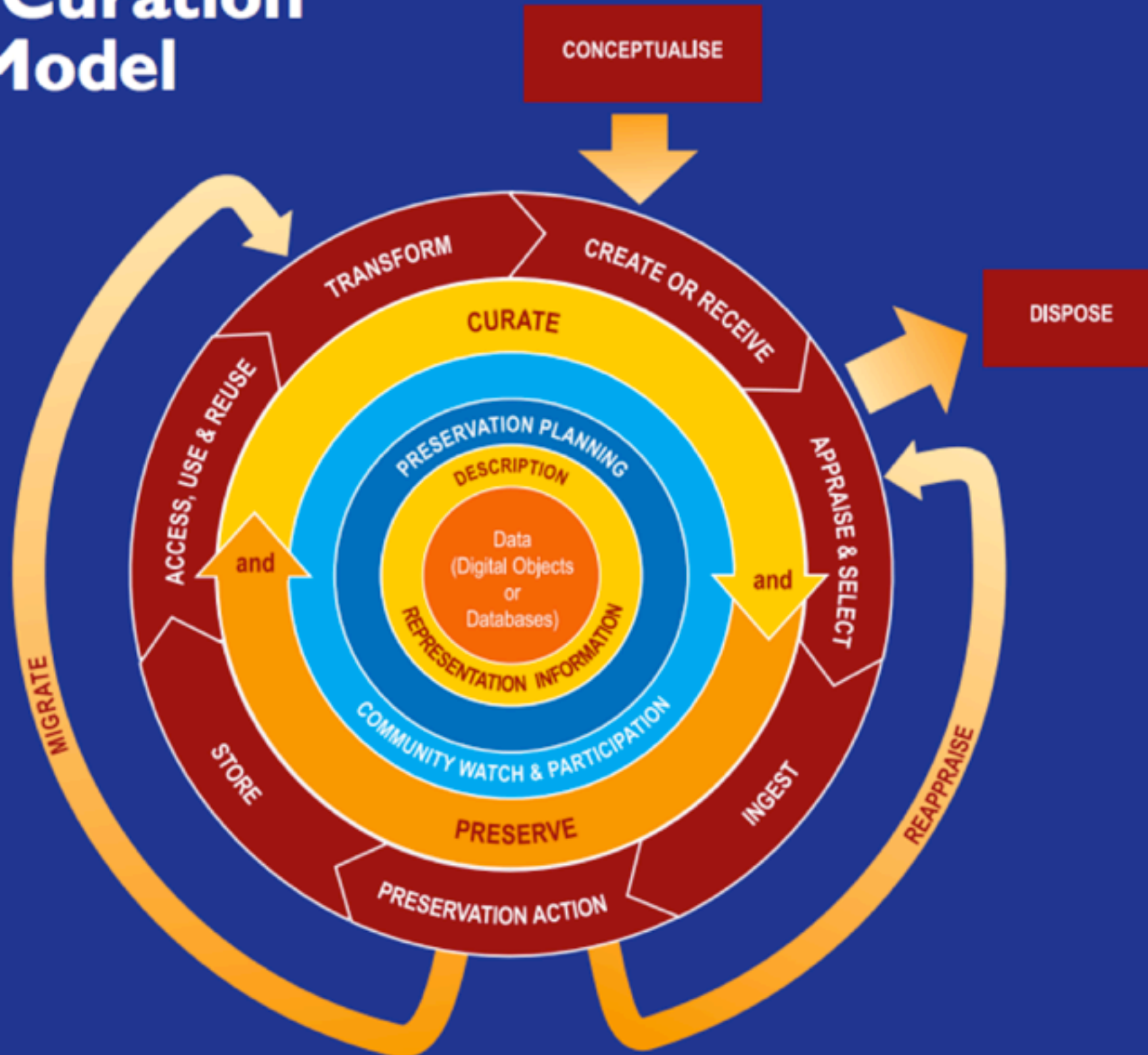




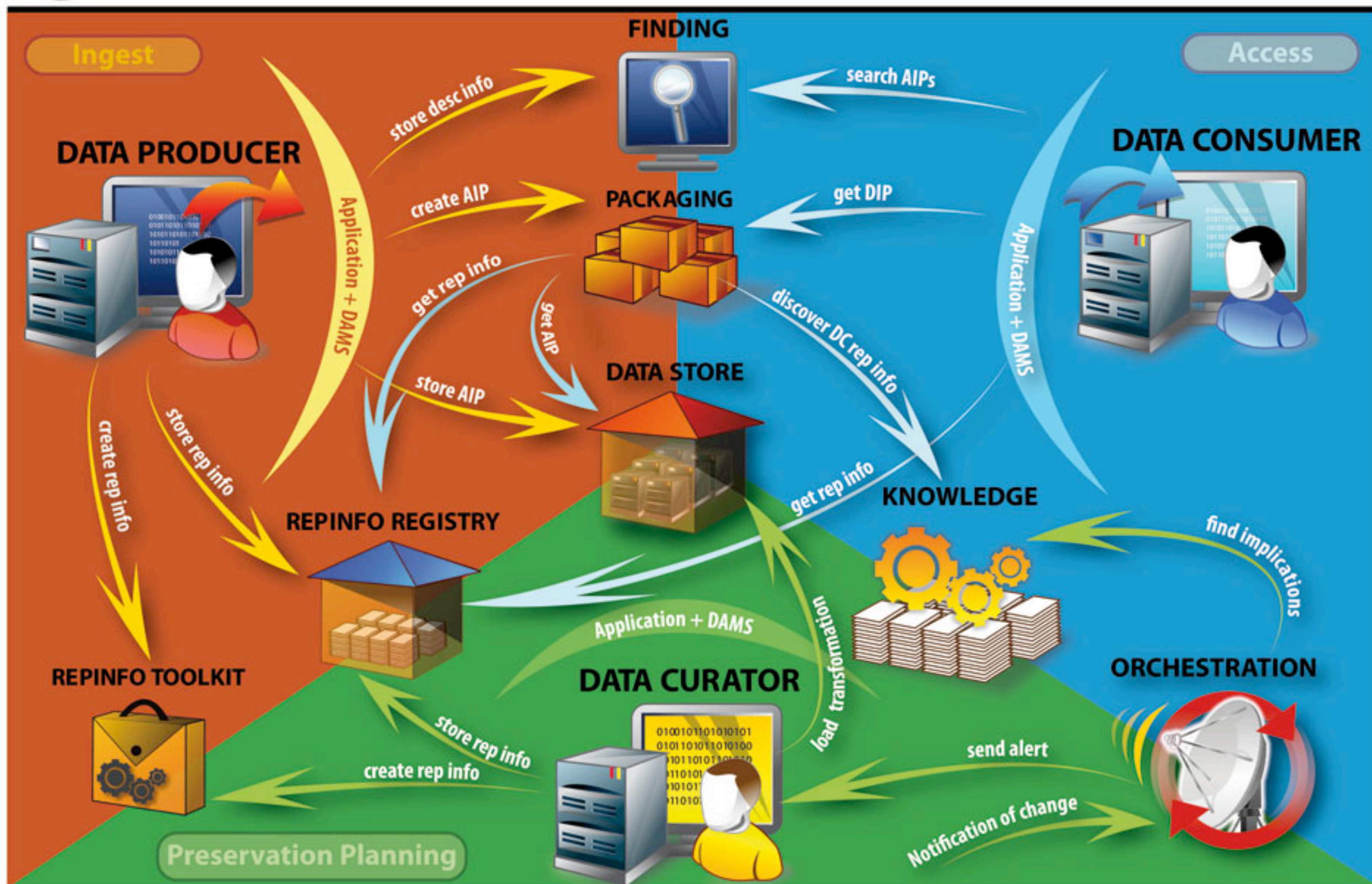
The DCC Curation Lifecycle Model



The DCC Curation Lifecycle Model



Monitorização
Informação de representação



O que é preservação digital?

Conjunto de actividades que visam assegurar o
acesso continuado à informação digital

...envolve pessoas, tecnologia e recursos...

Motivações



Sector público

Imposição legal
(e.g. **depósito legal**, incorporações obrigatórias)

Preservação do **património** e cultura

Transparência

Evidência de actividade pública



Sector privado

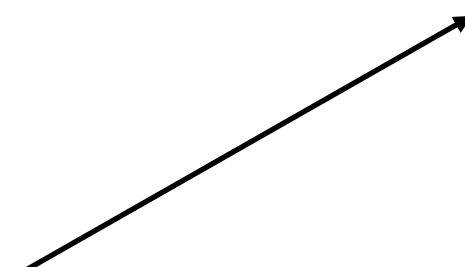
Imposição legal
(e.g. auditorias e **fiscalizações**, finanças, projectos financiados, etc.)

Razões **comerciais**/lucro

Protecção contra eventuais
litígios

Continuidade de negócio

O que muda essencialmente?



Central
Oklahoma's
Premiere
Shopper's Guide
Since 1977





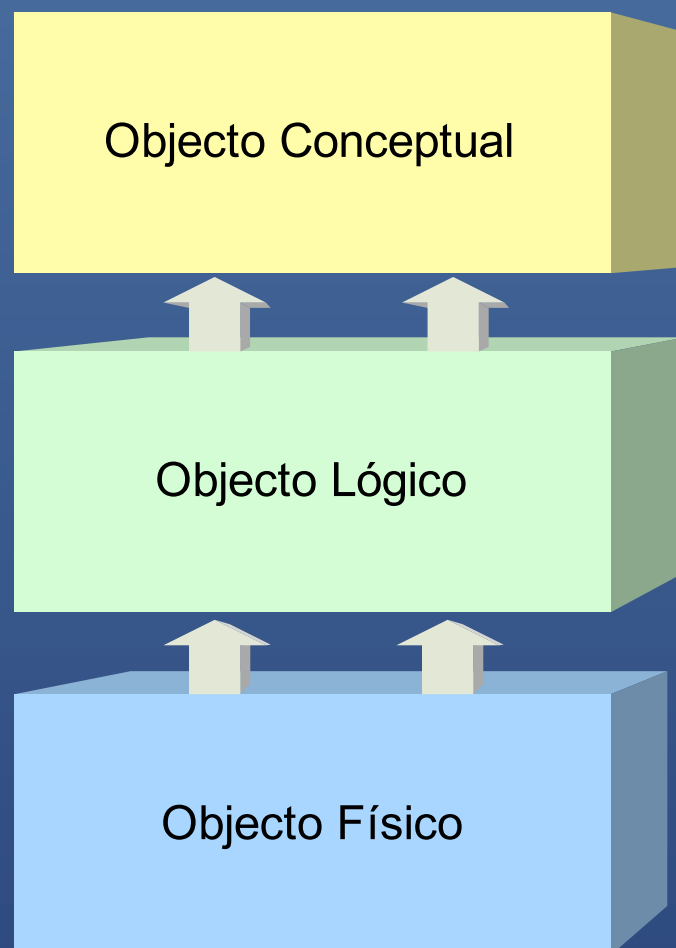












Nível Conceptual

Aquilo com que lidamos no **mundo real**, i.e. o que reconhecemos como uma **unidade tangível** de informação, e.g. jornal, livro, mapa, fotografia

Nível Lógico

Codificação da informação.

Tarefa realizada pelo software. Trata-se do formato do objecto

Nível Físico

Sinais físicos inscritos num suporte e.g. disco rígido, CD, Pen-drive

...se algum destes níveis se tornar inacessível perdemos
o acesso ao objecto conceptual...

...em menos de 5 anos, pelo menos um destes níveis será substituído devido a obsolescência...

5

tecnologias que se irão tornar obsoletas a breve prazo

5







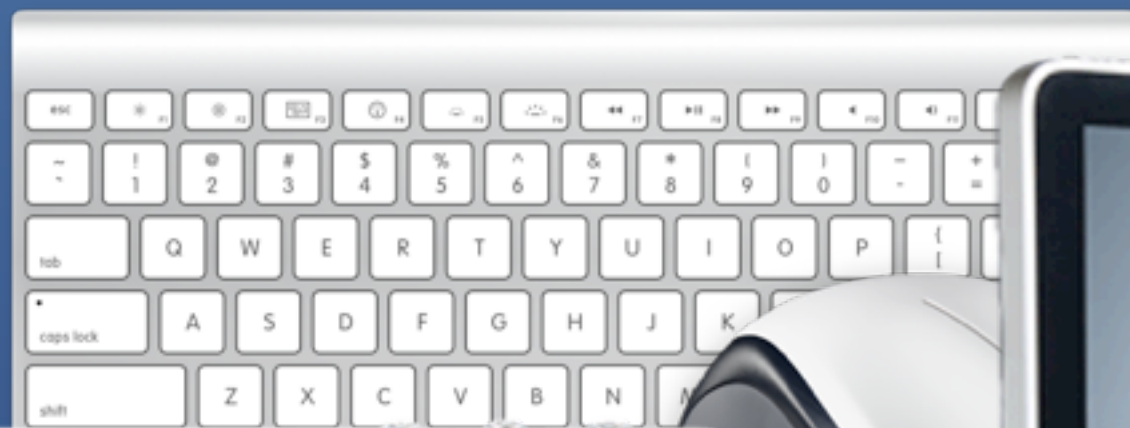


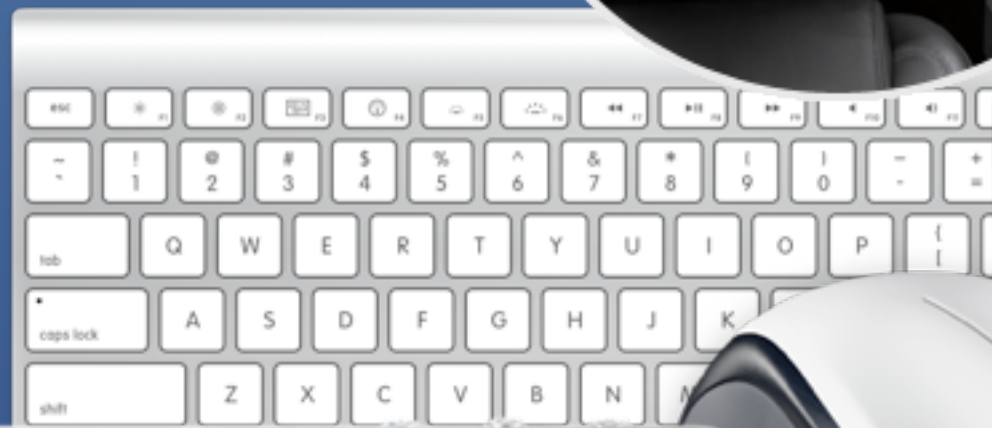
4











3







Dropbox





Dropbox



2







1



hulu



hulu



Spotify®

hulu



Spotify®

hulu



Spotify®



MEO VideoClube

O que podemos fazer?

Identificar claramente quais os **activos de informação**

Quais as tipologias de informação temos de preservar? Quais os formatos envolvidos? Qual o período de retenção de cada activo?

Identificar claramente quais os **activos de informação**

Quais as tipologias de informação temos de preservar? Quais os formatos envolvidos? Qual o período de retenção de cada activo?

Definir o **nível de preservação**

Que características dos objectos são fundamentais reter? Qual é o nosso SLA?

Identificar claramente quais os **activos de informação**

Quais as tipologias de informação temos de preservar? Quais os formatos envolvidos? Qual o período de retenção de cada activo?

Definir o **nível de preservação**

Que características dos objectos são fundamentais reter? Qual é o nosso SLA?

Definir um conjunto e **políticas**

Qual a missão do arquivo? Quais são as imposições legais que temos de respeitar?

Identificar claramente quais os **activos de informação**

Quais as tipologias de informação temos de preservar? Quais os formatos envolvidos? Qual o período de retenção de cada activo?

Definir o **nível de preservação**

Que características dos objectos são fundamentais reter? Qual é o nosso SLA?

Definir um conjunto e **políticas**

Qual a missão do arquivo? Quais são as imposições legais que temos de respeitar?

Planear

Analisar qual a melhor forma de fazer cumprir as políticas e assegurar o nível de preservação desejado. Que ferramentas temos ao nosso dispor? Testar várias alternativas.

Identificar claramente quais os **activos de informação**

Quais as tipologias de informação temos de preservar? Quais os formatos envolvidos? Qual o período de retenção de cada activo?

Definir o **nível de preservação**

Que características dos objectos são fundamentais reter? Qual é o nosso SLA?

Definir um conjunto e **políticas**

Qual a missão do arquivo? Quais são as imposições legais que temos de respeitar?

Planear

Analisar qual a melhor forma de fazer cumprir as políticas e assegurar o nível de preservação desejado. Que ferramentas temos ao nosso dispor? Testar várias alternativas.

Agir

Executar todas as operações necessárias à correcta aplicação das políticas definidas

Identificar claramente quais os **activos de informação**

Quais as tipologias de informação temos de preservar? Quais os formatos envolvidos? Qual o período de retenção de cada activo?

Definir o **nível de preservação**

Que características dos objectos são fundamentais reter? Qual é o nosso SLA?

Definir um conjunto e **políticas**

Qual a missão do arquivo? Quais são as imposições legais que temos de respeitar?

Planear

Analisar qual a melhor forma de fazer cumprir as políticas e assegurar o nível de preservação desejado. Que ferramentas temos ao nosso dispor? Testar várias alternativas.

Agir

Executar todas as operações necessárias à correcta aplicação das políticas definidas

Controlar a **qualidade**

As acções que estamos a aplicar estão a produzir os resultados esperados?

Identificar claramente quais os **activos de informação**

Quais as tipologias de informação temos de preservar? Quais os formatos envolvidos? Qual o período de retenção de cada activo?

Definir o **nível de preservação**

Que características dos objectos são fundamentais reter? Qual é o nosso SLA?

Definir um conjunto e **políticas**

Qual a missão do arquivo? Quais são as imposições legais que temos de respeitar?

Planear

Analisar qual a melhor forma de fazer cumprir as políticas e assegurar o nível de preservação desejado. Que ferramentas temos ao nosso dispor? Testar várias alternativas.

Agir

Executar todas as operações necessárias à correcta aplicação das políticas definidas

Controlar a **qualidade**

As acções que estamos a aplicar estão a produzir os resultados esperados?

Monitorizar

Conhecer os riscos internos e externos à organização por forma a reagir atempadamente a novas ameaças

Identificar claramente quais os **activos de informação**

Quais as tipologias de informação temos de preservar? Quais os formatos envolvidos? Qual o período de retenção de cada activo?

Definir o **nível de preservação**

Que características dos objectos são fundamentais reter? Qual é o nosso SLA?

Definir um conjunto e **políticas**

Qual a missão do arquivo? Quais são as imposições legais que temos de respeitar?

Planear

Analisar qual a melhor forma de fazer cumprir as políticas e assegurar o nível de preservação desejado. Que ferramentas temos ao nosso dispor? Testar várias alternativas.

Agir

Executar todas as operações necessárias à correcta aplicação das políticas definidas

Controlar a **qualidade**

As acções que estamos a aplicar estão a produzir os resultados esperados?

Monitorizar

Conhecer os riscos internos e externos à organização por forma a reagir atempadamente a novas ameaças

Auditar

Aferir através de entidades isentas se os nossos procedimentos estão a ser executados de acordo com o estabelecido

Dados

Repositórios digitais

Conteúdo da Web

Dados científicos

Dados

Repositórios digitais

Conteúdo da Web

Dados científicos



Caracterização

(Identificação de formatos, extracção de propriedades)

Dados

Repositórios digitais

Conteúdo da Web

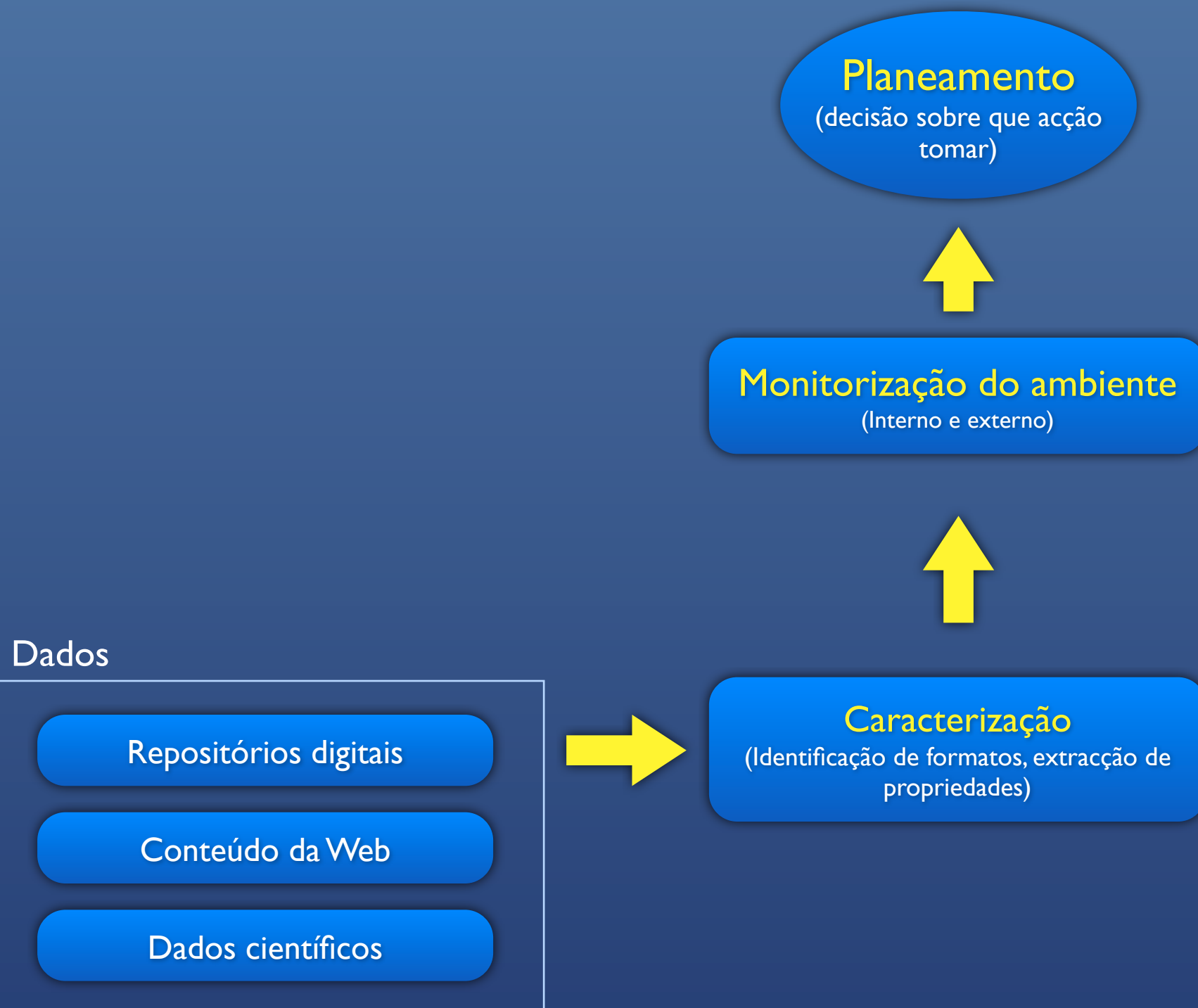
Dados científicos

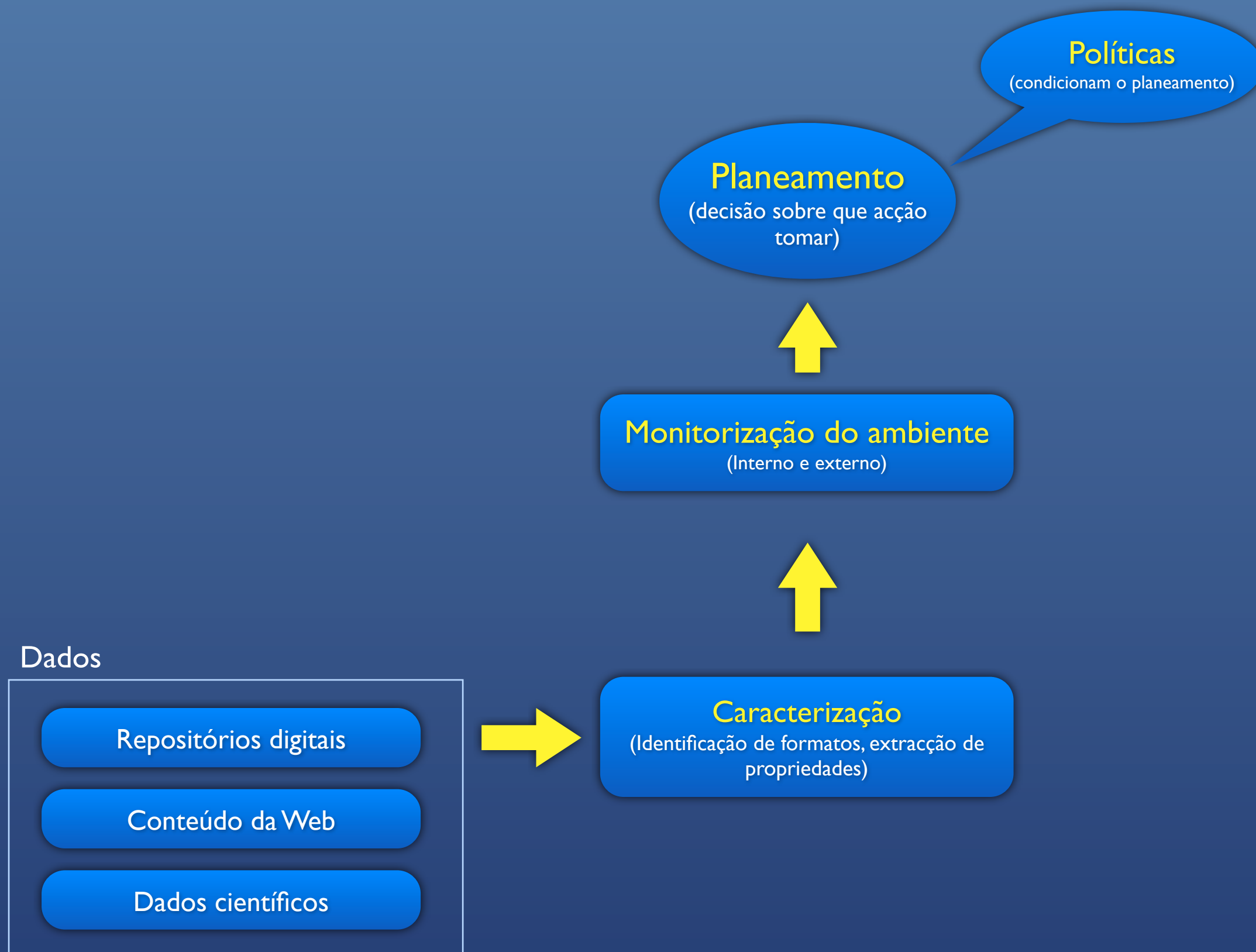


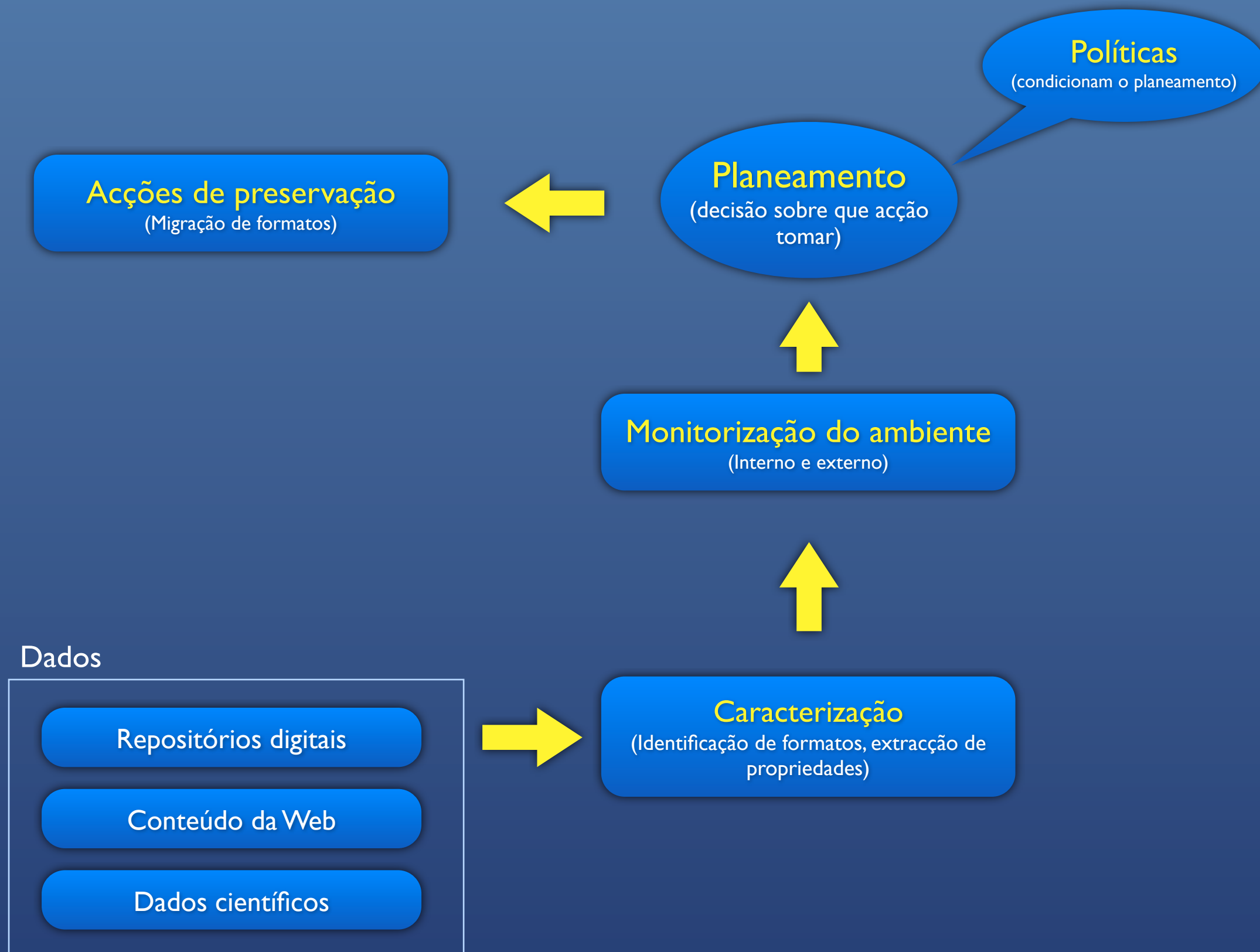
Monitorização do ambiente
(Interno e externo)

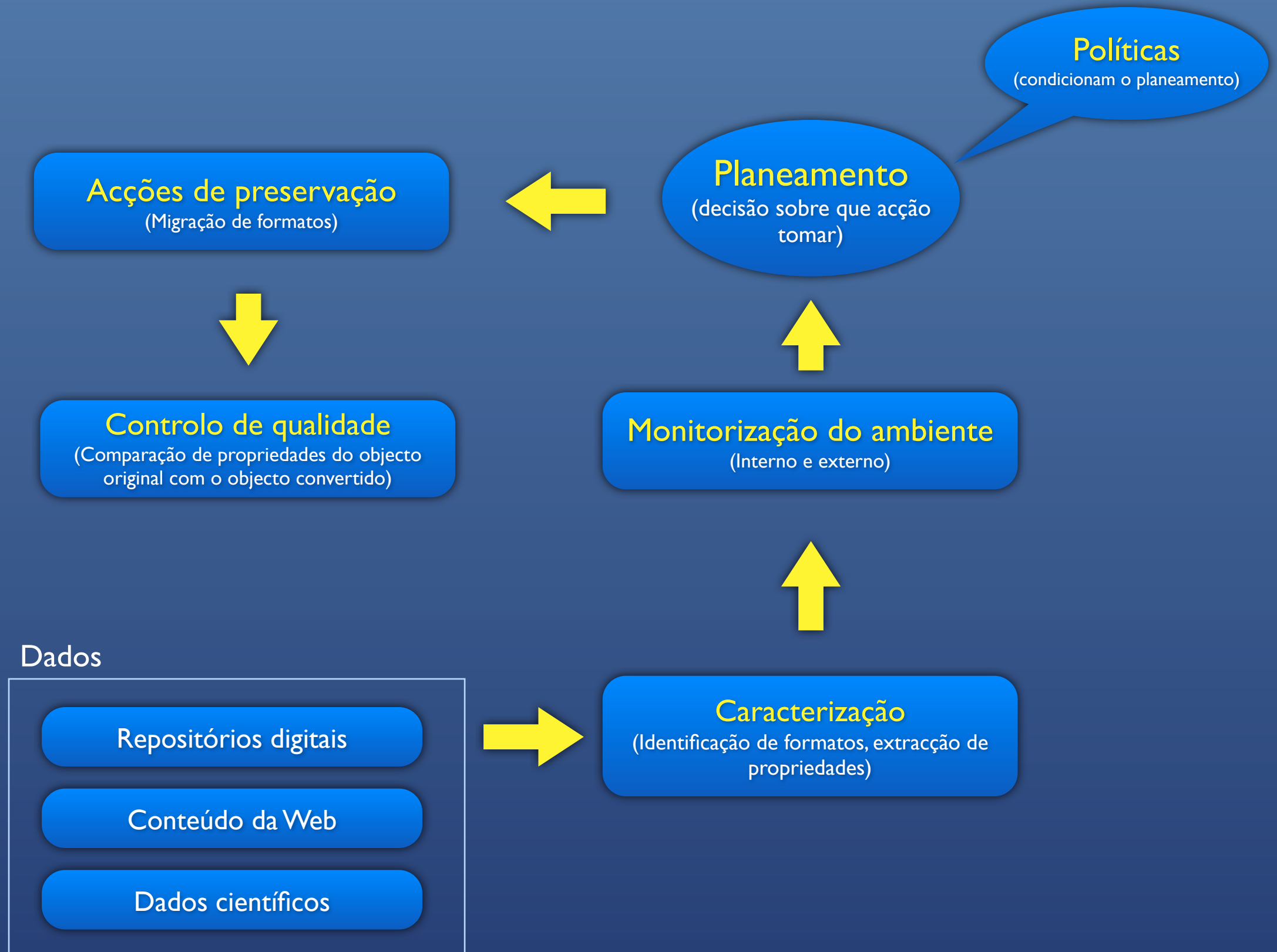


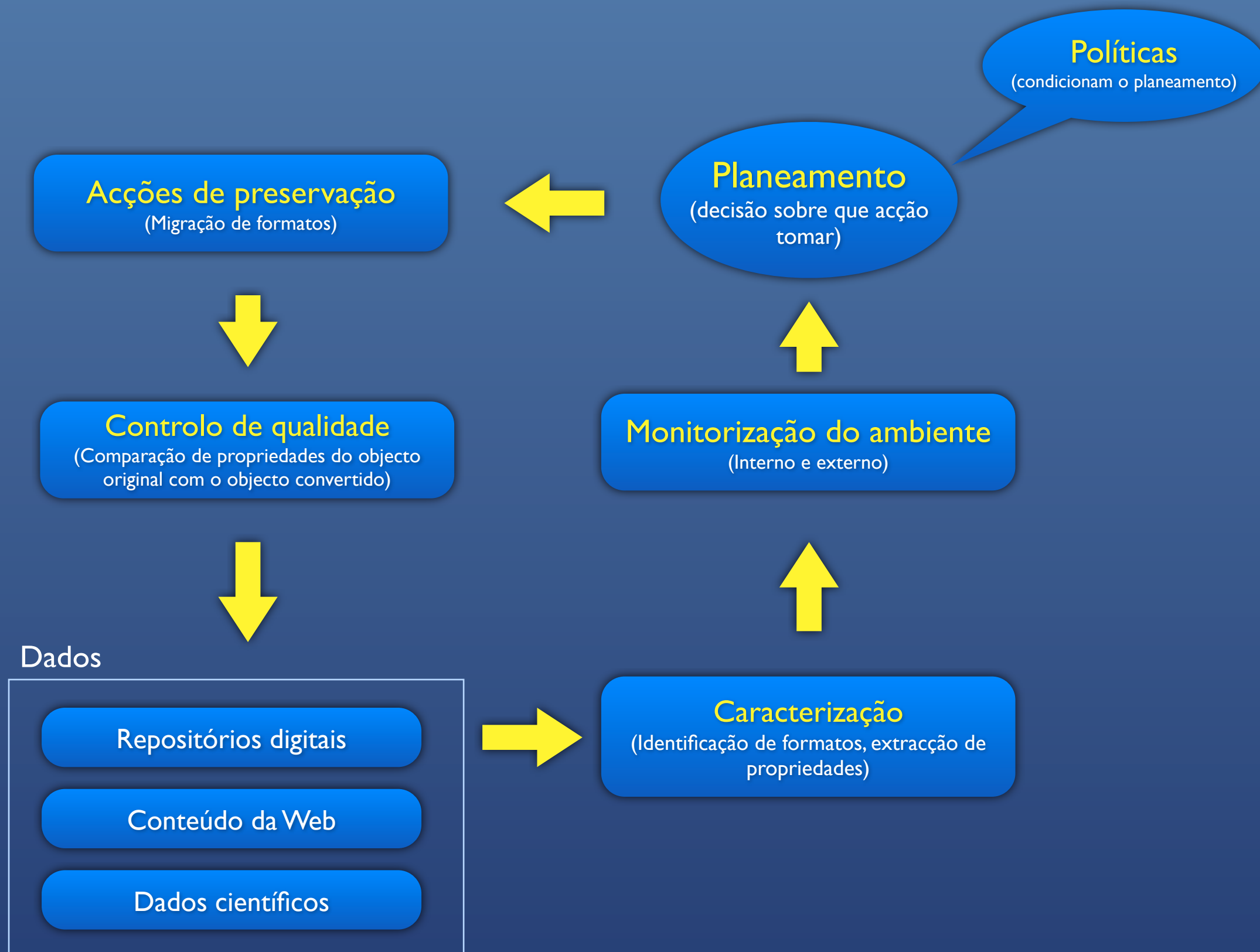
Caracterização
(Identificação de formatos, extracção de propriedades)







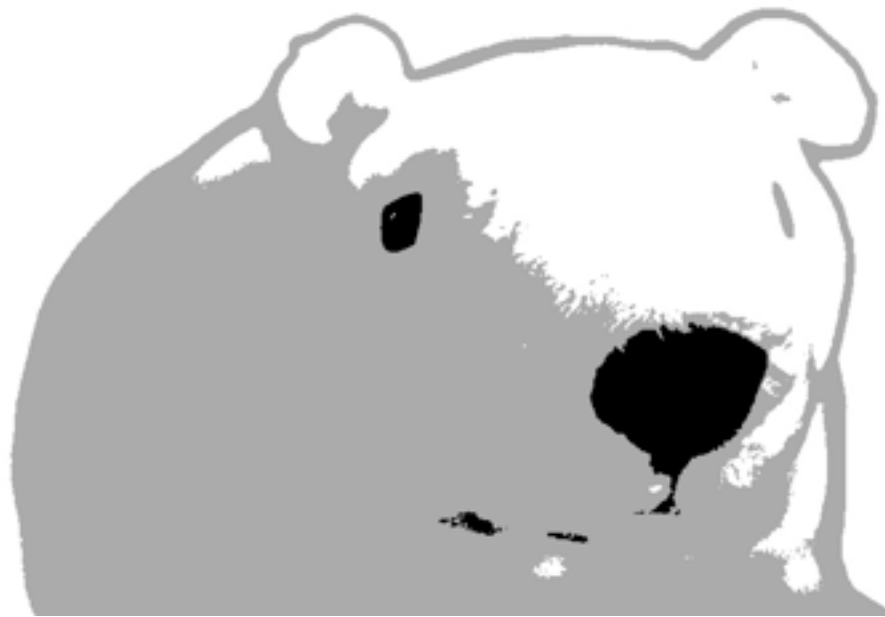




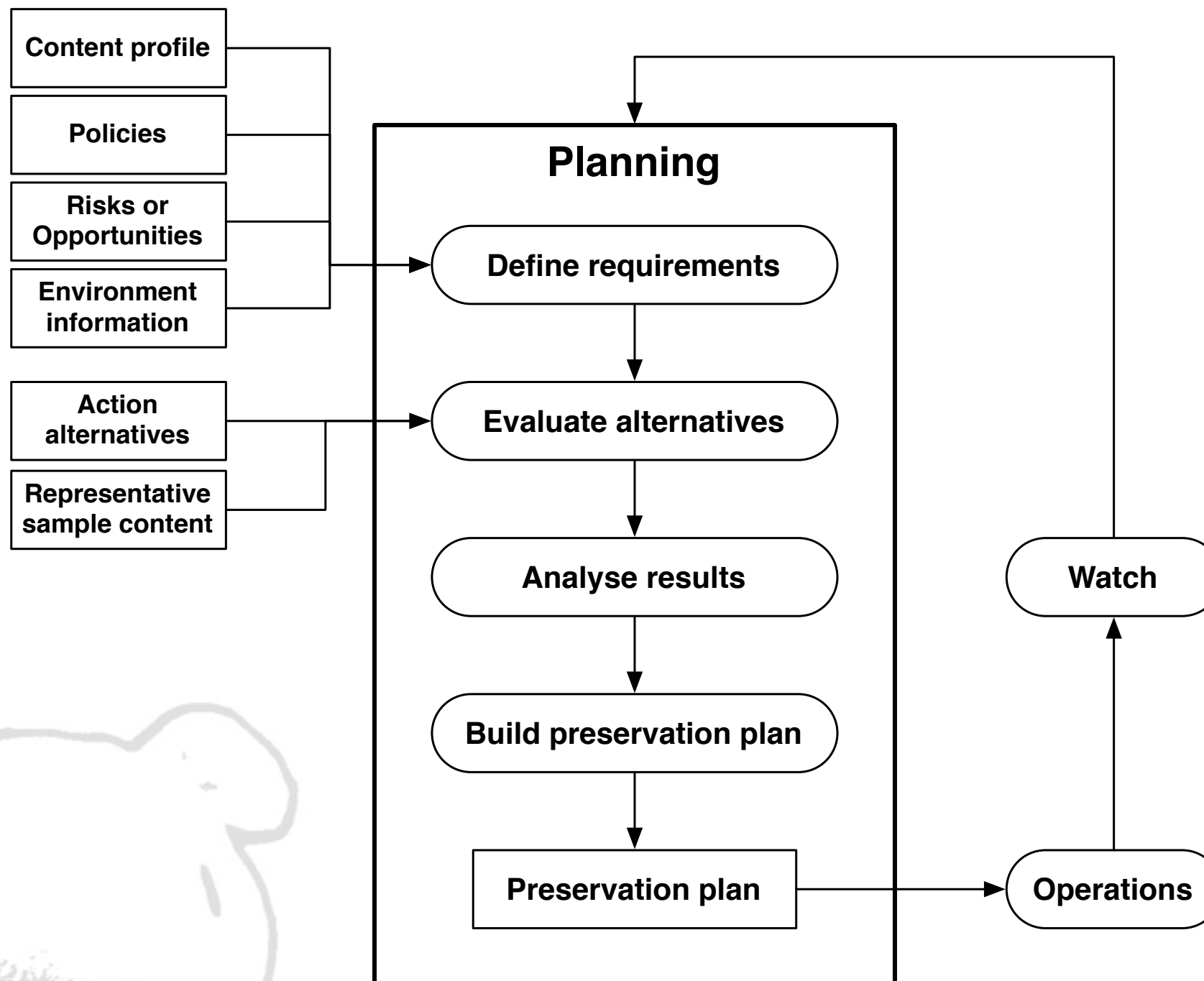
Como planejar?

PLATO

Preservation Planning Tool



<http://www.ifs.tuwien.ac.at/dp/plato>



Como agir?

Antes de ingerir

Manualmente impondo políticas

Automaticamente recorrendo a software

Depois de ingerir

Recorrendo a software especializado ou a funcionalidade disponível no software de gestão da informação

Ao aceder

Mantendo a informação conforme foi ingerida e apenas fazer algo no momento em que esta for solicitada

Como controlar a qualidade?

Objecto A



Propriedades de A



Objecto A'



Propriedades de A'

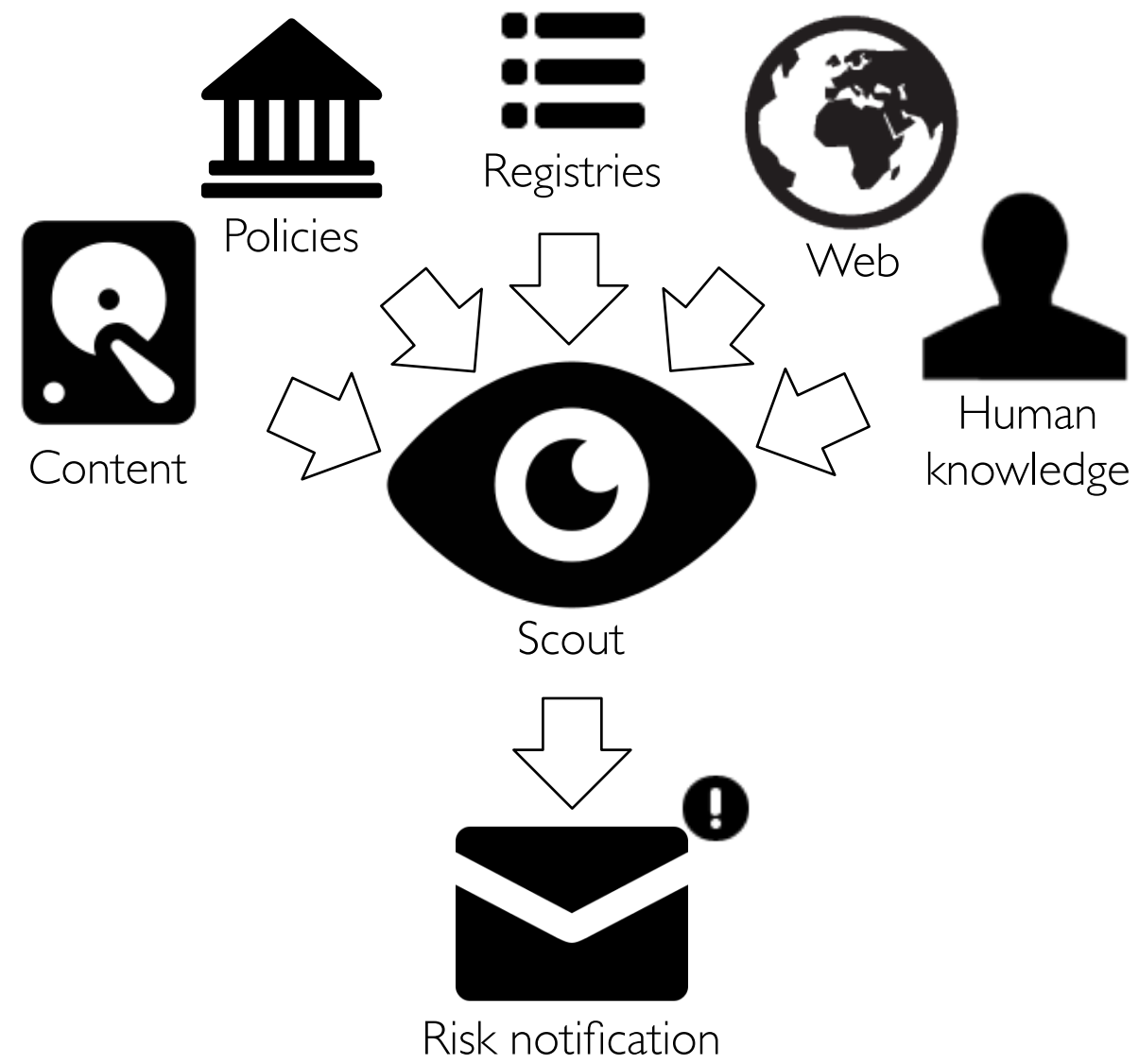


Controlo de qualidade
e.g. xCorrSound, fprobe, Matchbox, MarcAlizer, Jpylyzer



Similaridade

Como monitorizar?



Como auditar?

ISO/DIS 16363

Audit and certification of trustworthy digital repositories

DRAMBORA

Digital Repository Audit Method Based on Risk Assessment

DSA

Data Seal of Approval

O que está actualmente na agenda de investigação europeia?



Como desenvolver **soluções escaláveis**?

Como processar milhões de objectos, centenas de formatos, ficheiros na ordem dos gigabytes em tempo útil?

Como **quantificar os custos** de preservação?

Como transformar a preservação num modelo de negócio e não apenas num custo?

Como **integrar sistemas** de forma natural e transparente?

Como integrar sistemas de arquivo corrente com sistemas de preservação digital?

Como desenvolver **soluções escaláveis**?

Como processar milhões de objectos, centenas de formatos, ficheiros na ordem dos gigabytes em tempo útil?

Como **quantificar os custos** de preservação?

Como transformar a preservação num modelo de negócio e não apenas num custo?

Como **integrar sistemas** de forma natural e transparente?

Como integrar sistemas de arquivo corrente com sistemas de preservação digital?



1

notícia
em segunda mão



RODA community
REPOSITORY OF AUTHENTIC DIGITAL OBJECTS


[RODA](#)
[DOWNLOAD](#)
[SUPPORT](#)
[COMMUNITY](#)
[CONTACT](#)

RODA

The world's most advanced **open-source** digital repository



An open-source **digital repository** designed for preservation

RODA is a complete digital repository that delivers functionality for all the main units of the OAIS reference model. RODA is capable of ingesting, managing and providing access to the various types of digital content produced by large corporations or public bodies. RODA is based on open-source technologies and is supported by existing standards such as the OAIS, METS, EAD and PREMIS.



Conforms to open standards

RODA follows open standards using EAD for description metadata, PREMIS for preservation metadata, METS for structural metadata, several standards for technical metadata (e.g. NISO Z39.87 for digital still images).



Vendor independent

RODA is 100% built on top of open-source technologies. The entire infrastructure required to support RODA is vendor independent. This means that you may use the hardware and Linux distributions that best fit your institutional needs.



Scalable

The service-oriented nature of RODA's architecture allows the system to be highly



Embedded preservation actions

Preservation actions and management within RODA is handled by a task scheduler. The task

Download | Repository of Authentic Digital Objects

www.roda-community.org/download/

Férias design ret redmine KEEP roda Save to Mendeley crm e-ark SAZ coolK adprt weebox@seb 4C SCAPE RODA WP confluence arch Issues

RODA community
REPOSITORY OF AUTHENTIC DIGITAL OBJECTS

RODA DOWNLOAD SUPPORT COMMUNITY CONTACT

Download

RODA can be experienced in many different ways. You may download the source-code straight from GitHub and compiling it for maximum performance and adaptation to your platform, download the binary package for easy installation in your own server, or download the ready-to-use Virtual Machine for testing purposes.

You may also try the [online demo](#) before downloading RODA.

Virtual Machine



To try out RODA without installing it, you can use a RODA virtual machine. To use it you'll need to do the following:

1. Download VirtualBox from [here](#) (version 4.2.10 or higher) and install it on your local machine
2. Download RODA virtual machine from [here](#) (Right click > Save link as...)

4.2GB, SHA: 6ad1004ddb0420e422ace0852cb23f8c07eb9e43

1. Start VirtualBox, go to menu File -> Import Appliance... and select the file RODA.ova that you just downloaded
2. When import finishes, start the virtual machine and wait 1 or 2 minutes for the machine to boot
3. Open your browser and go to <http://localhost:8080/> to access RODA Web User Interface. You can login using username **admin** and password **roda**.

To access the virtual machine console, use an SSH client and connect to **localhost:9922** using login **roda** and password **roda**.

You can change the ports 8080 (HTTP) and 9922 (SSH) going to RODA virtual machine Settings > Network > Adapter 1 > Advanced > Port Forwarding.

Governance

Open-source projects are not typically governed by a central entity or group of people. They usually evolve in an uncontrolled form driven by the common wishes of its community. However, in order to build such a community, core people must exist in order for the project to start, prosper and grow.

In the particular case of RODA we feel that is necessary to establish a stable network of people and organisations that will engage interested parties to join and participate in its development. In order to accomplish this goal, we propose the existence of a governance structure that takes responsibility for various aspects of the project. This structure is bound to evolve as the community matures and grows.

Executive board

The executive board is composed of a small group of people responsible for establishing the mission and the vision of the project. The approval of the roadmap for each major version of the software is one of the main responsibilities of this board.

The board is currently composed by the following people:



Miguel Ferreira

CEO of KEEP SOLUTIONS, Fellow researcher at the Algoritmi Research Centre of the University of Minho

Ph.D. by the University of Minho in Technologies and Information Systems with a specialization in Digital Preservation. Degree in Systems Engineering and Computer Science by the University of Minho. He is a founding member of KEEP SOLUTIONS. Prior to joining KEEP SOLUTIONS, has worked as a researcher at the University of Minho, consultant at the OPorto Regional Archives and as a researcher for Philips Research.



Francisco Barbedo

Head of department for innovation and e-government of the Portuguese National Archives

Former Deputy Director of the National Archives, has a Masters in Information Management and a Degree in History of Art and Archeology. Francisco was the project manager of the first RODA release and was an



KEEP SOLUTIONS is a company that provides advanced services for managing and preserving digital information. KEEP SOLUTIONS sponsored the infrastructure that supports RODA Community. This includes hosting costs, design, documentation, man-power and community support. KEEP SOLUTIONS has also co-funded the european project SCAPE that has further enhanced RODA to support scalable preservation actions and planning.

Website: <http://www.keep.pt>

Period: Continuous ongoing support since 2008



SCAPE – Scalable Preservation Environments – is a european research project funded under the 7th Framework Programme of the European Commission that aims at developing scalable services for planning and execution of institutional preservation strategies on an open source platform that orchestrates semi-automated workflows for large-scale, heterogeneous collections of complex digital objects. RODA has been enhanced to support scalable preservation actions and planning under FP7 ICT-2009.4.1 call and Grant Agreement number 270137. The work on RODA within this project was also co-funded by KEEP SOLUTIONS.

Website: <http://www.scape-project.eu>

Period: 2011-2014



The original funders of RODA were the POAP (Operational Programme for Public Administration), a national financial instrument funded by the European Union (75%), and the Directorate-General of Portuguese Archives (25%). The POAP aimed at organisational innovation, streamlining administrative procedures, improving public services quality and training staff in central administration.

Website: <http://www.dgarq.gov.pt>

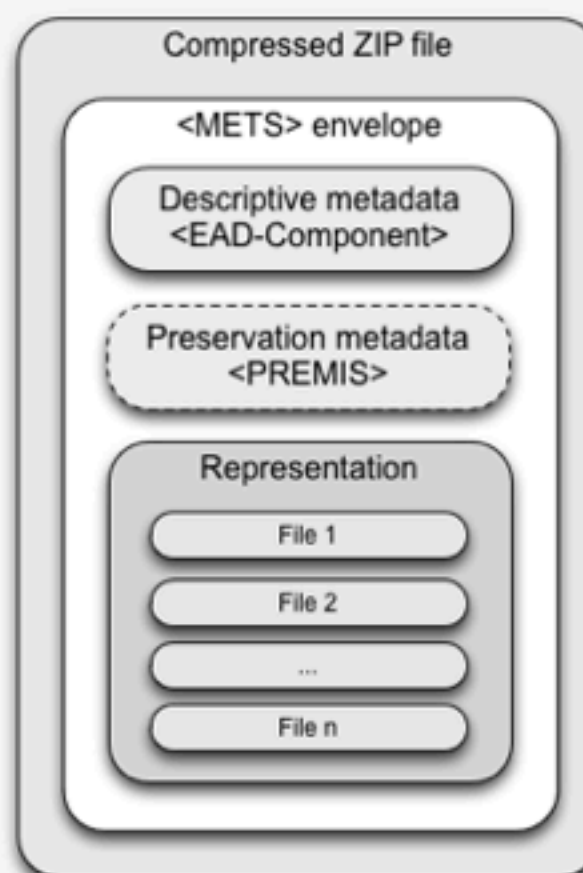
Period: 2006-2008

Submission Information Packages (SIP)

New entries to the repository come in the shape of Submission Information Packages (SIP). When the ingest process terminates, SIPs are transformed into **Archival Information Packages (AIP)**, i.e. the actual packages that will be kept in the repository. Associated with the AIP is the structural, technical and preservation metadata, as they are essential for carrying out preservation activities.

The SIP is composed of one or more digital representations and all of the associated metadata, packaged inside a METS envelope. Producers take advantage of a small application called RODA-in that allows them to create these packages. The structure of a SIP supported by RODA is depicted in the following Figure. The RODA SIP is basically a compressed ZIP file containing a METS envelope, the set of files that compose the representations and a series of metadata records. Within the SIP there should be at least one descriptive metadata record in EAD-Component format.

One may also find preservation and technical metadata inside a submission package, although this last set of metadata is not mandatory as is seldom created by producers. Nevertheless, it was felt important that RODA should support those additional SIP elements for special situations such as repository succession, e.g. when ingested items belong to a repository that is to be deactivated.



[Developer guide](#)[Publications](#)[Mailing lists](#)[Bug tracker](#)[Paid support](#)

License

RODA is licensed under [LGPLv3](#) for all source-code including interoperability libraries like SIP manipulation libraries. A full listing of the license is provided next:

GNU LESSER GENERAL PUBLIC LICENSE

Version 3, 29 June 2007

Copyright © 2007 Free Software Foundation, Inc.

Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed.

This version of the GNU Lesser General Public License incorporates the terms and conditions of version 3 of the GNU General Public License, supplemented by the additional permissions listed below.

0. Additional Definitions.

As used herein, “this License” refers to version 3 of the GNU Lesser General Public License, and the “GNU GPL” refers to version 3 of the GNU General Public License.

“The Library” refers to a covered work governed by this License, other than an Application or a Combined Work as defined below.

An “Application” is any work that makes use of an interface provided by the Library, but which is not otherwise based on the Library. Defining a subclass of a class defined by the Library is deemed a mode of using an interface provided by the Library.

Developer guide · keeps/roda Wiki

GitHub, Inc.

github.com/keeps/roda/wiki/Developer-guide

Reader

Férias ▾ design ▾ ret redmine ▾ KEEP ▾ roda crm e-ark SAZ coolK adprt weebox@seb 4C ▾ SCAPE RODA WP confluence arch Issues SCAPE-Overview Koha ▾ sms Sacro ▾ >> +

Home Pages History

New Page

Edit Page Page History Clone URL

Developer guide

This guide will have information on:

1. [How to get the source code](#)
2. [RODA build system \(Apache Maven\)](#)
3. [How the code is laid out](#)
4. [How to build and run](#)
5. [How to set up the development environment](#)
6. [How to contribute](#)

How to get the source code

You can easily get the source code by cloning the project into your machine (just needs git installed):

```
$ git clone https://github.com/keeps/roda.git roda.git
```

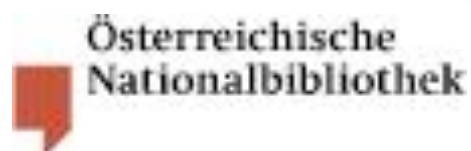
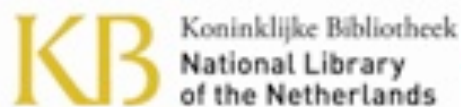
If you plan to contribute to RODA, you will need to first fork the repository into your own GitHub account and then clone it into your machine. To learn how to do it, please check this [GitHub article](#).

RODA build system (Apache Maven)

RODA uses [Apache Maven](#) build system. Being a multi-module Maven project, in the root **pom.xml** is declared all the important information to all modules in RODA, such as:

- Modules to be included in the default build cycle
- Maven repositories to be used
- Dependency management (version numbers are declared here and inherited by the sub-modules)
- Plugin management (version numbers are declared here and inherited by the sub-modules)
- Profiles available (e.g., RODA installer is **ONLY** created when the profile **create-installer** is used, and if you are developing and you want

www.roda-community.org



Technische Universität Berlin



Hoiame mõtteid, et jagada teiega

EESTI RAHVUSRAAMATUKOGU
NATIONAL LIBRARY OF ESTONIA



University of Essex



DET KONGELIGE BIBLIOTEK
DANMARKS NATIONALBIBLIOTEK OG KØBENHAVNS UNIVERSITETSbibliOTEK





www.keep.pt

Miguel Ferreira
mferreira@keep.pt