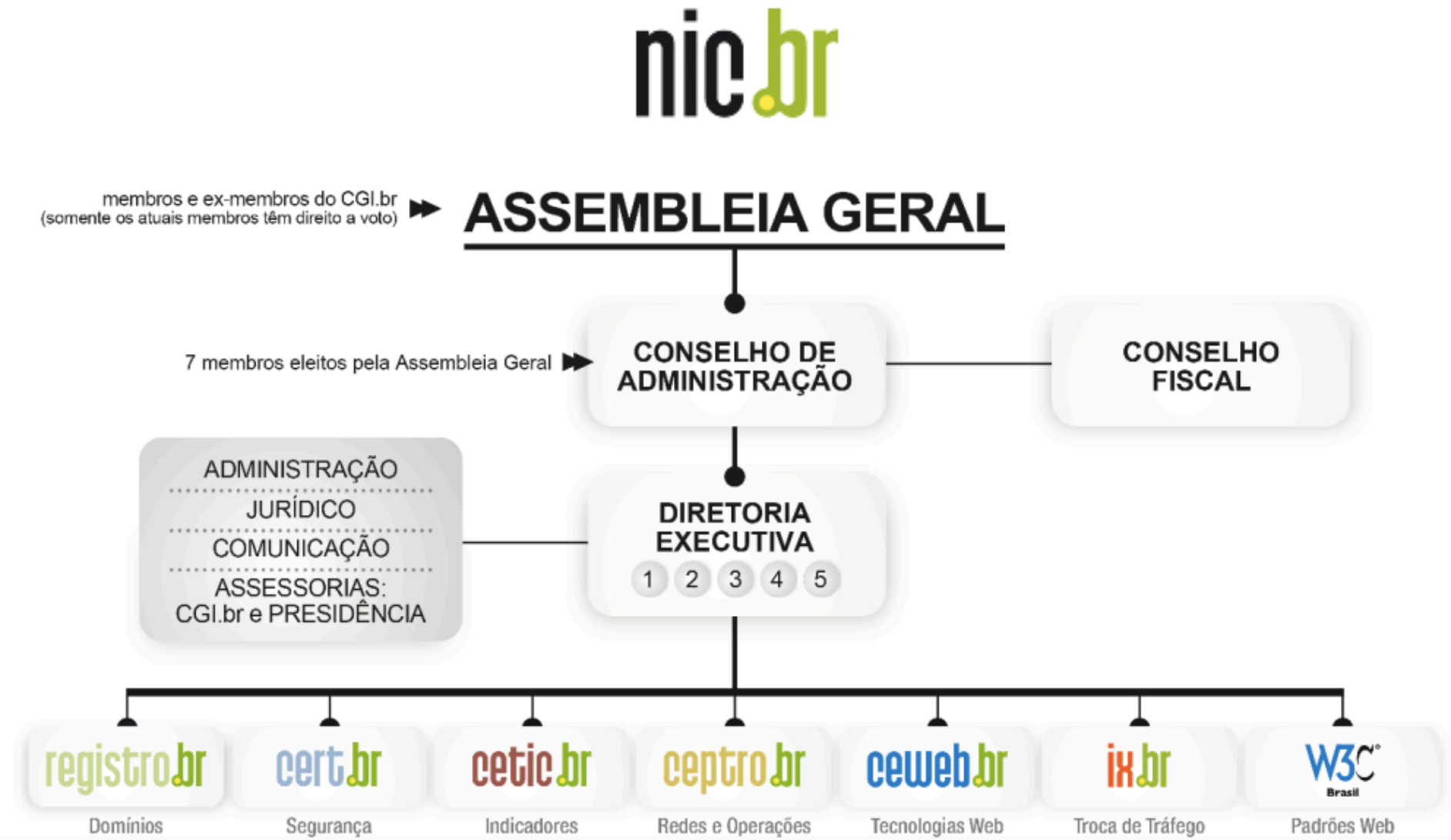
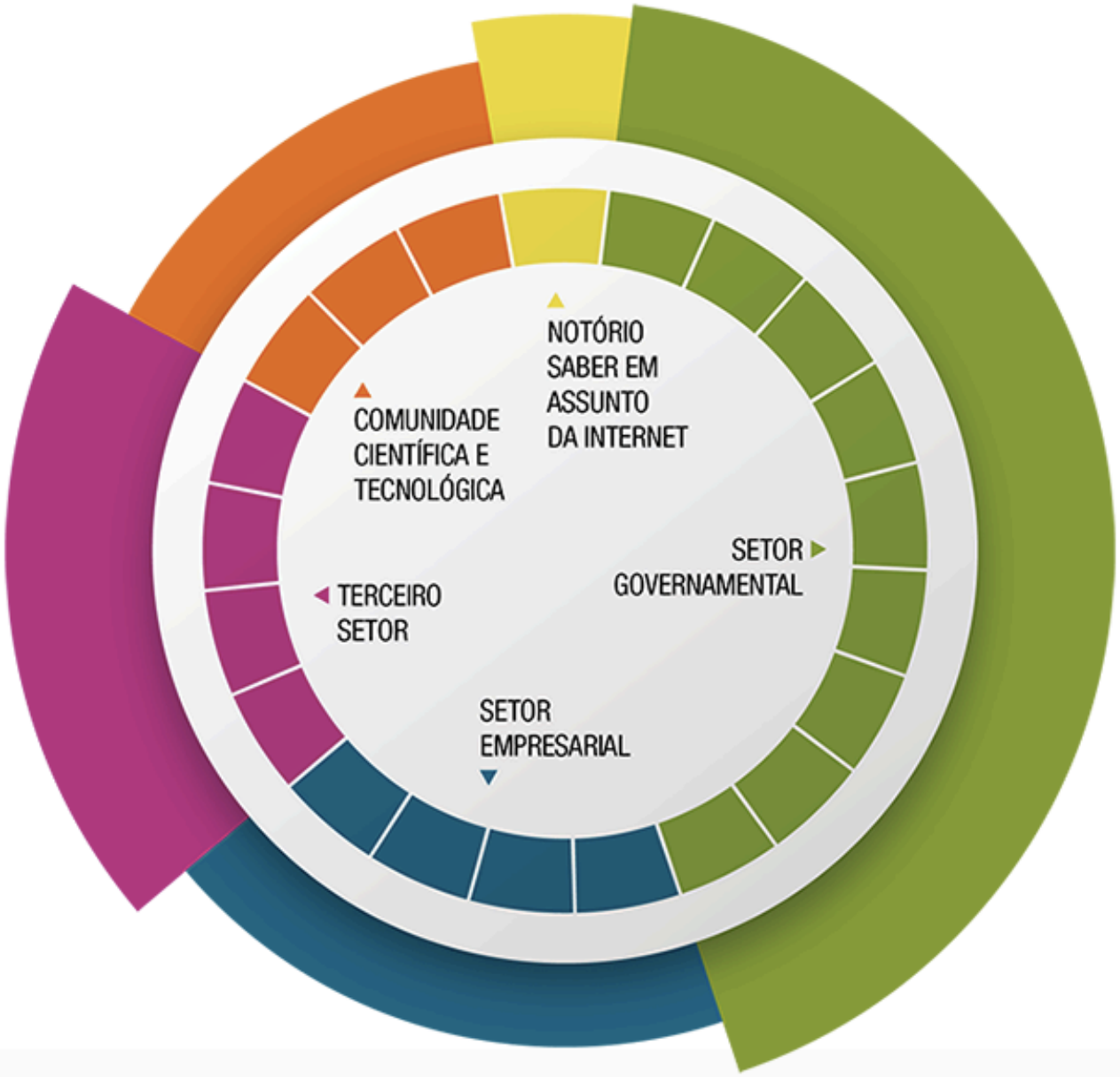
Buenas Prácticas para Datos en la Web: Desafíos y Beneficios

Bernadette Lóscio, Caroline Burle y Newton Calegari

Condatos, 25 de agosto de 2017

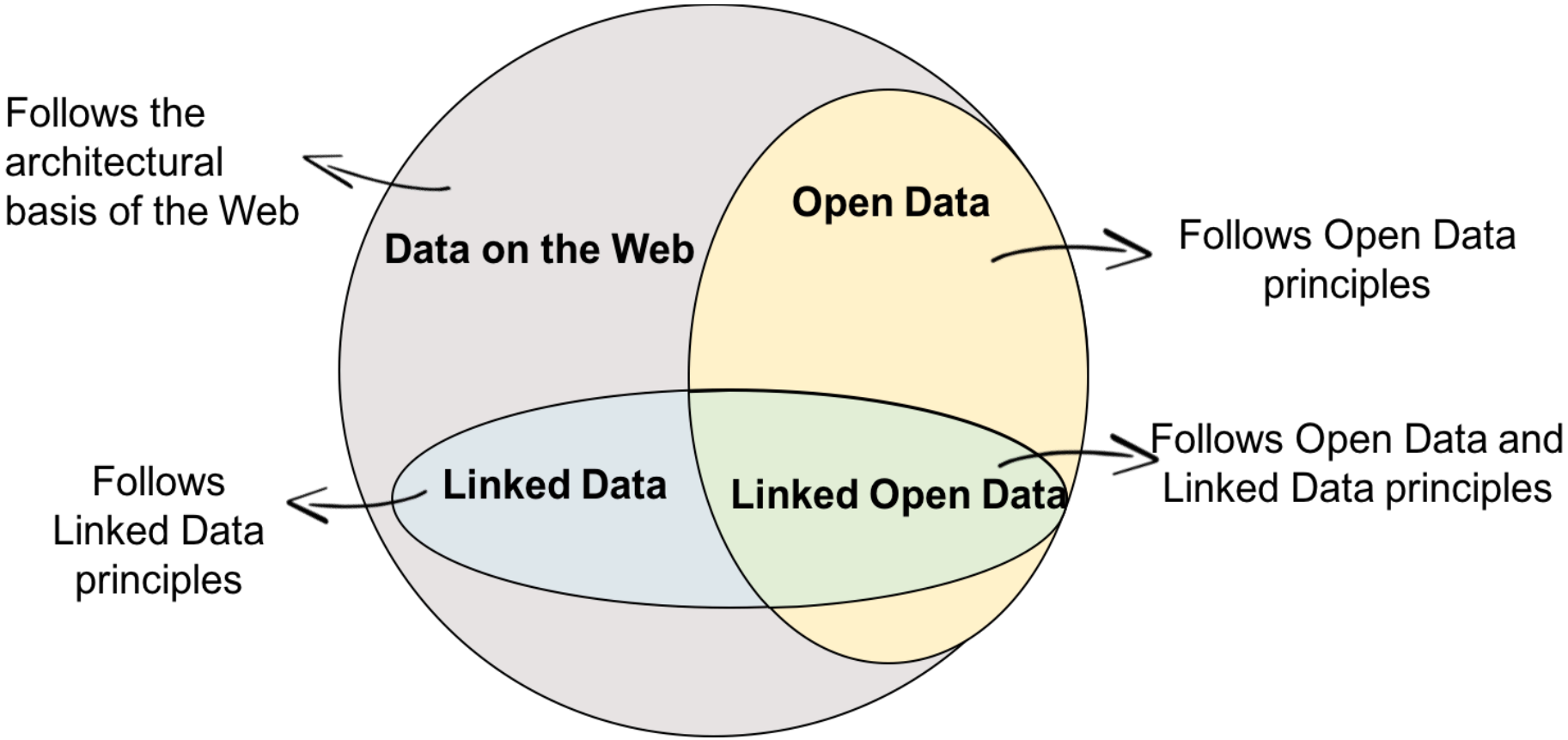

nic.br
membros e ex-membros do CGI.br
(somente os atuais membros têm direito a voto)
ASSEMBLEIA GERAL
7 membros eleitos pela Assembleia Geral
CONSELHO DE ADMINISTRAÇÃO
CONSELHO FISCAL
ADMINISTRAÇÃO
JURÍDICO
COMUNICAÇÃO
ASSESSORIAS:
CGI.br e PRESIDÊNCIA
DIRETORIA EXECUTIVA
1 2 3 4 5
registro.br
Domínios
cert.br
Segurança
cetic.br
Indicadores
ceptro.br
Redes e Operações
ceweb.br
Tecnologias Web
ix.br
Troca de Tráfego
W3C Brasil
Padrões Web


NOTÓRIO
SABER EM
ASSUNTO
DA INTERNET
COMUNIDADE
CIENTÍFICA E
TECNOLÓGICA
SETOR
GOVERNAMENTAL
TERCEIRO
SETOR
SETOR
EMPRESARIAL

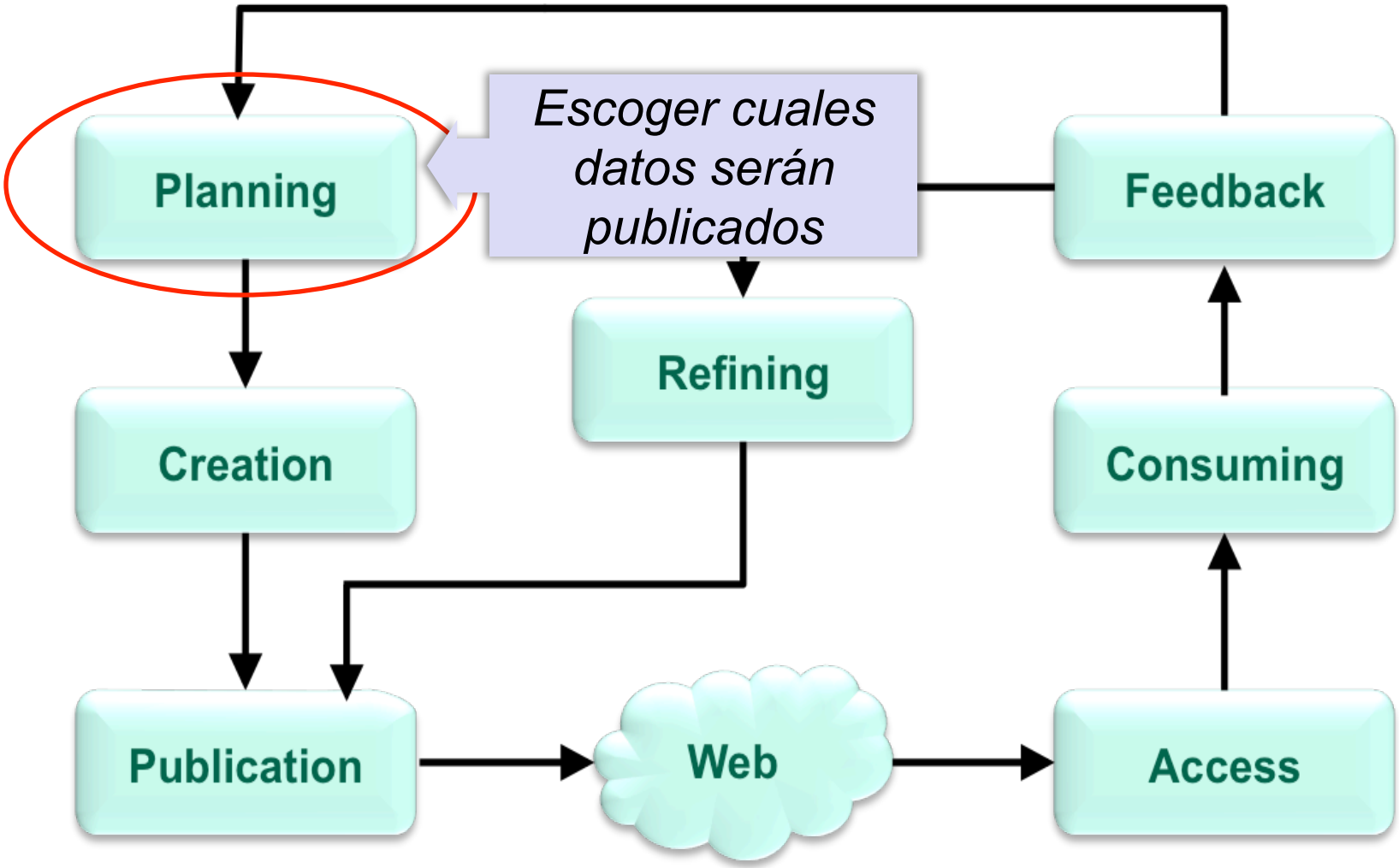
Tópicos que serán discutidos

- Contexto de la Web de Datos
- Casos de Uso de Datos en la Web
- Desafíos y Requisitos de Datos en la Web
- Buenas Prácticas de Datos en la Web
- Beneficios de las Buenas Prácticas de Datos en la Web

Datos en la Web x Datos Abiertos x Datos Conectados

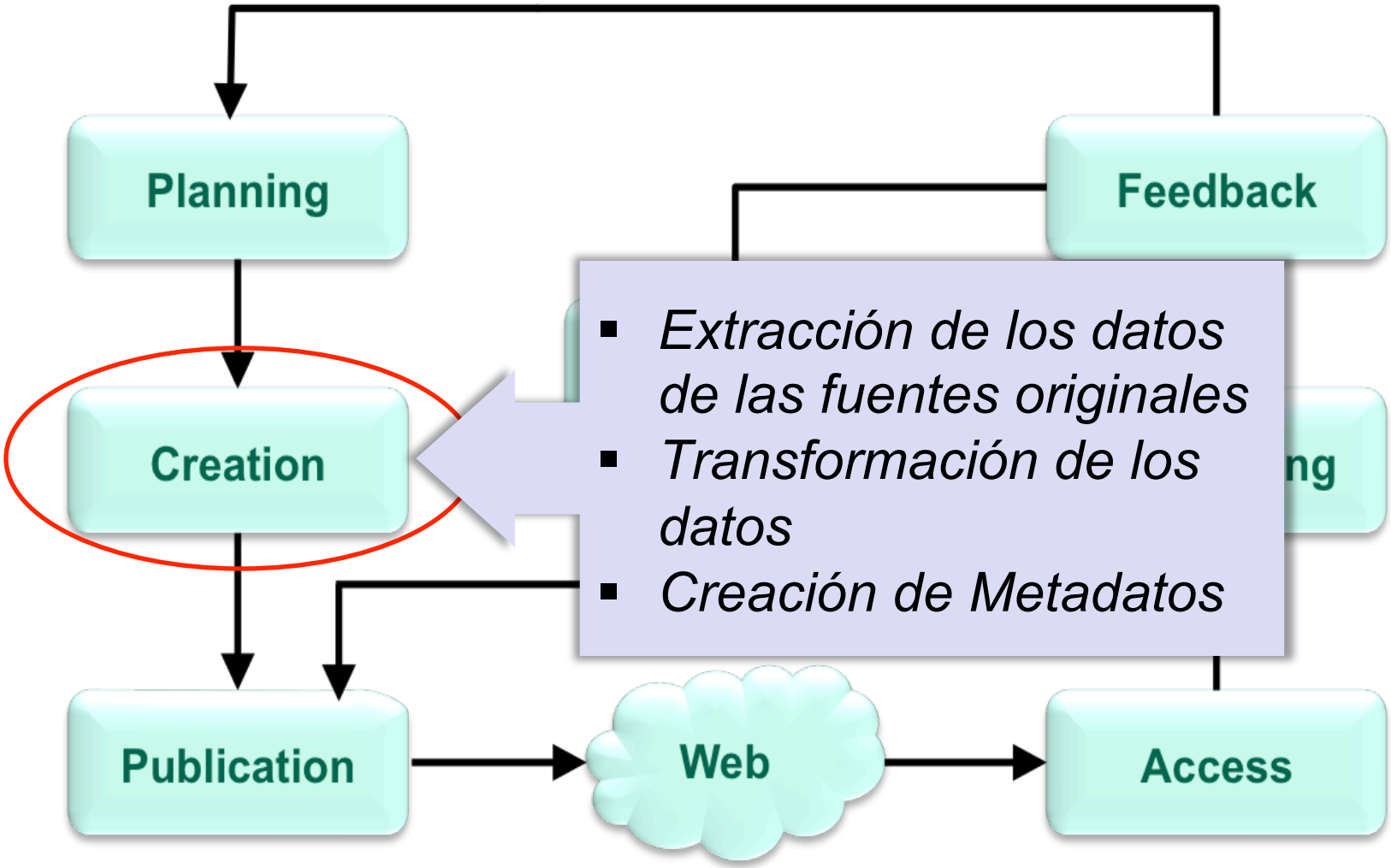


Ciclo de Vida de los Datos en la Web



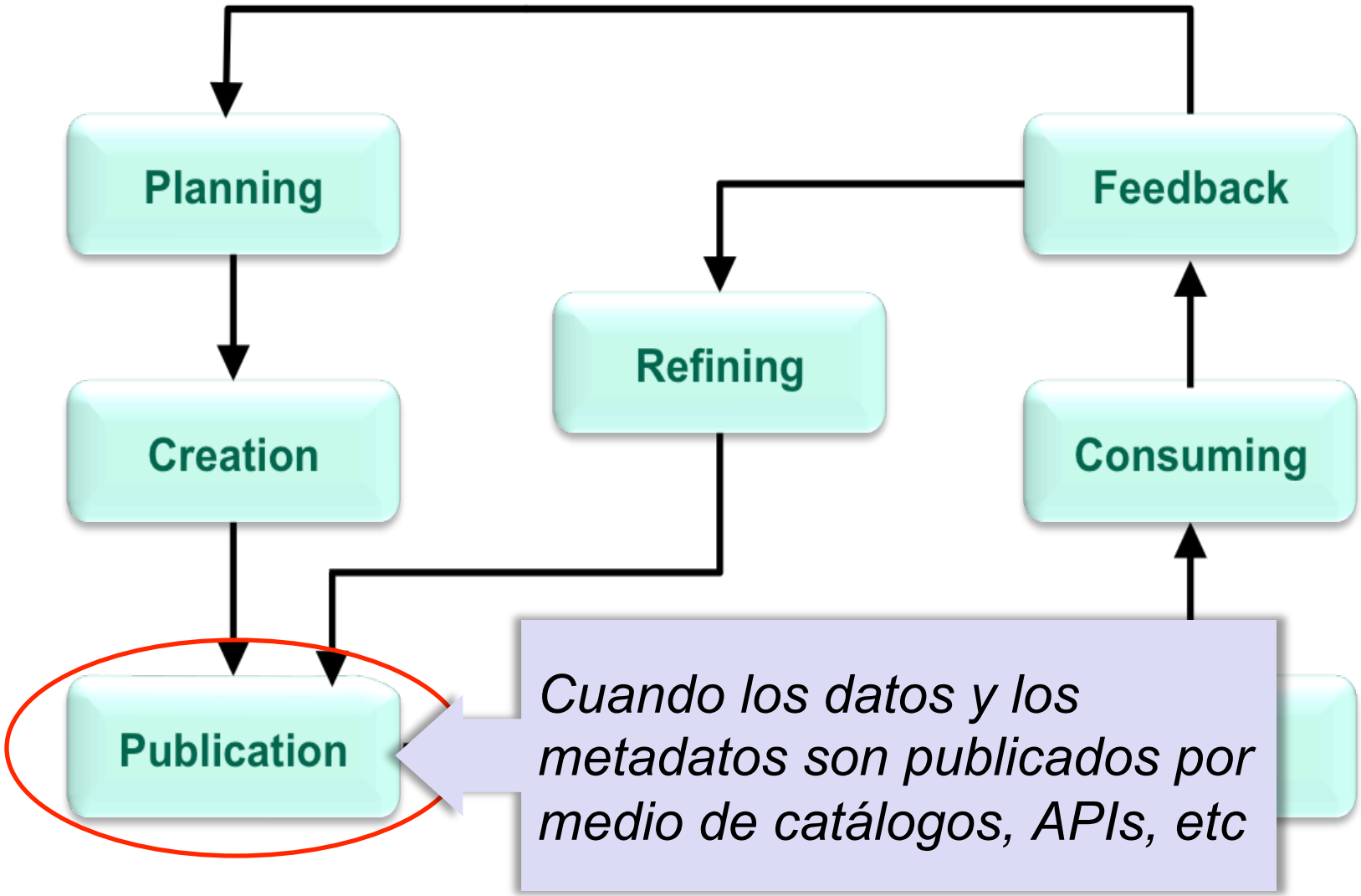
Fonte: http://www.slideshare.net/carolineburle/data-on-the-web-big-data-and-open-data

Ciclo de Vida de los Datos en la Web



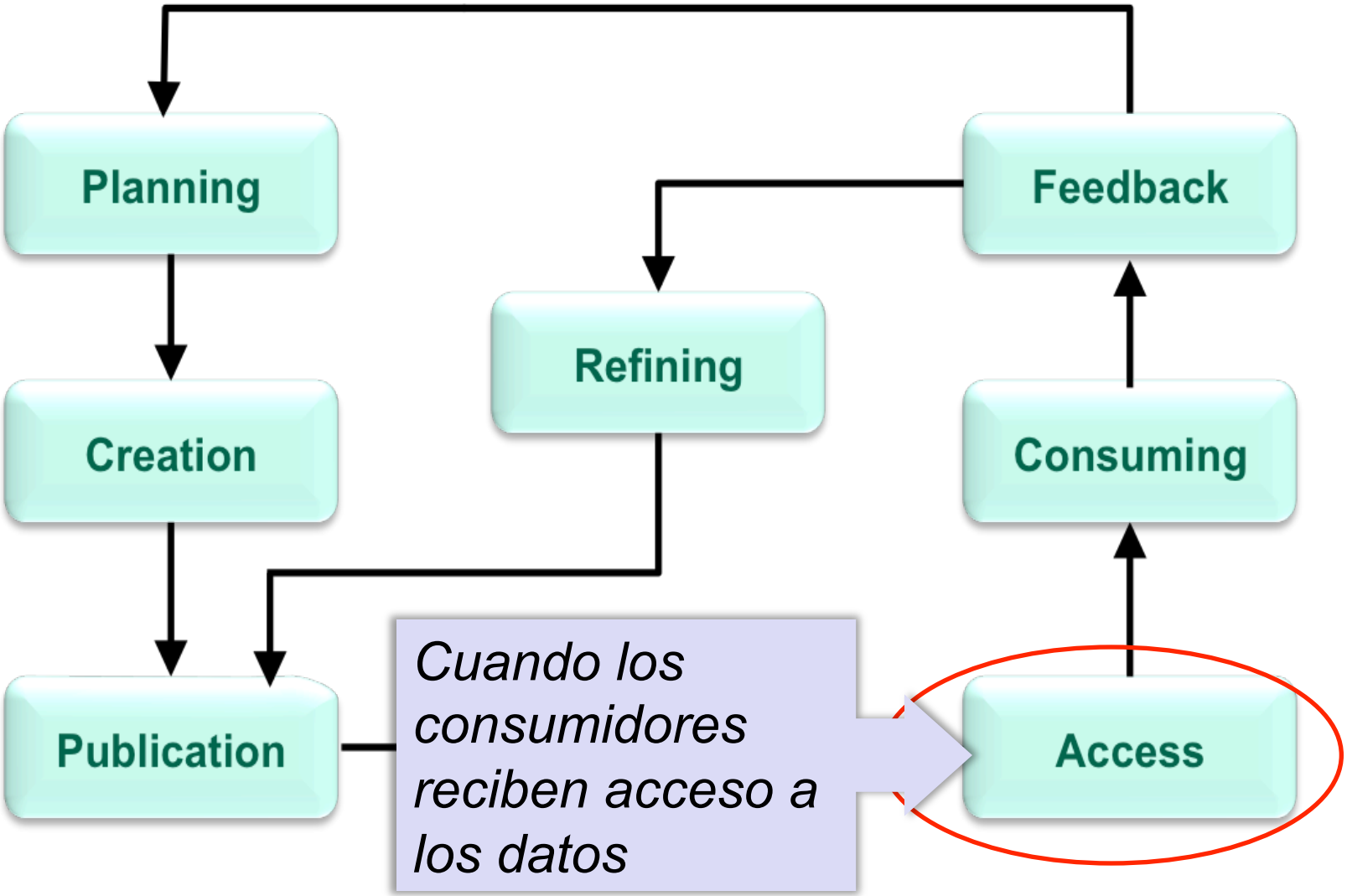
Fonte: http://www.slideshare.net/carolineburle/data-on-the-web-big-data-and-open-data

Ciclo de Vida de los Datos en la Web



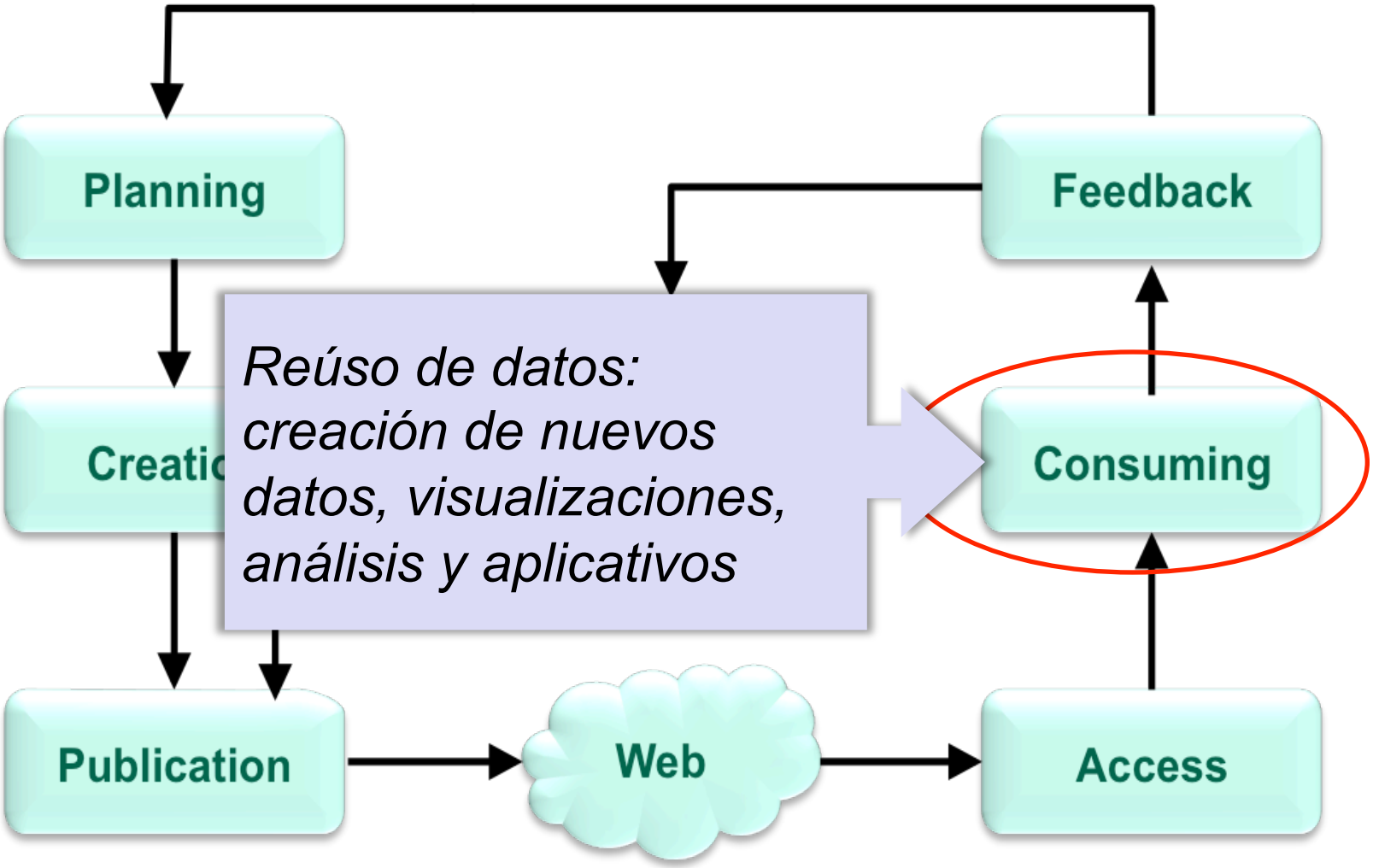
Fonte: http://www.slideshare.net/carolineburle/data-on-the-web-big-data-and-open-data

Ciclo de Vida de los Datos en la Web



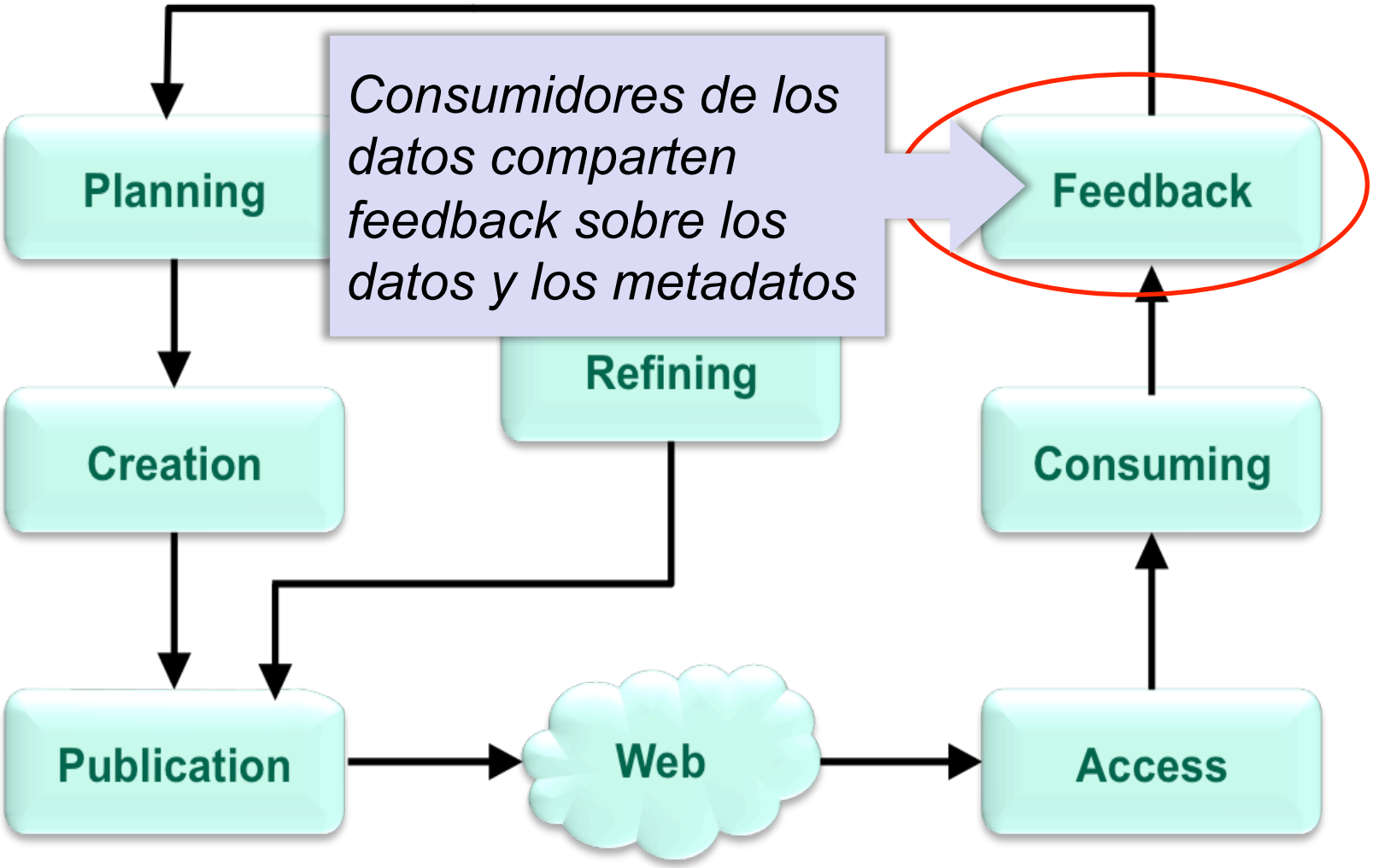
Fuente: http://www.slideshare.net/carolineburle/data-on-the-web-big-data-and-open-data

Ciclo de Vida de los Datos en la Web



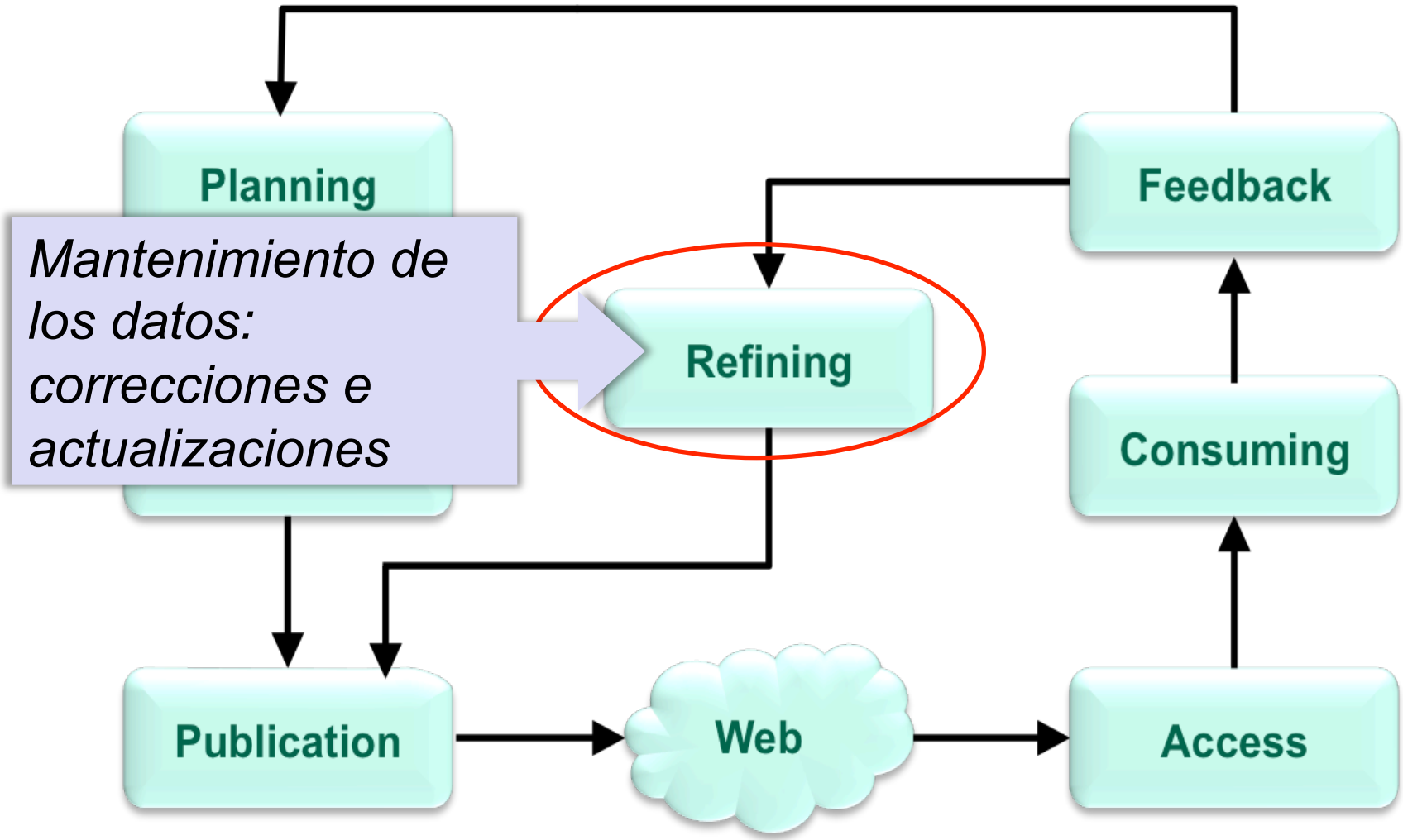
Fonte: http://www.slideshare.net/carolineburle/data-on-the-web-big-data-and-open-data

Ciclo de Vida de los Datos en la Web



Fonte: http://www.slideshare.net/carolineburle/data-on-the-web-big-data-and-open-data

Ciclo de Vida de los Datos en la Web



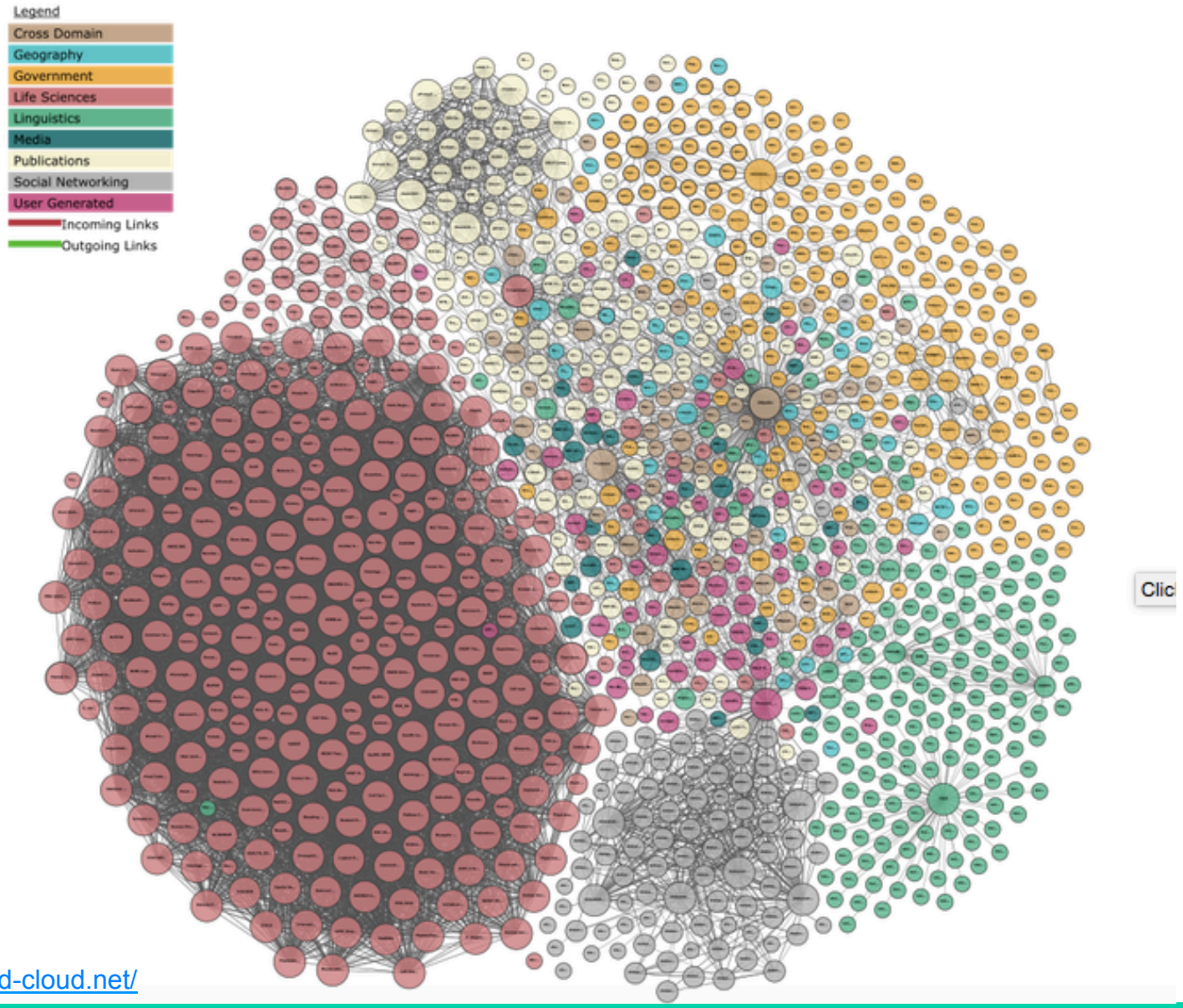
Fonte: http://www.slideshare.net/carolineburle/data-on-the-web-big-data-and-open-data

5 Estrellas de los Datos Abiertos

Mientras más estrellas se alcancen, más fácil se ubican y reutilizan los datos

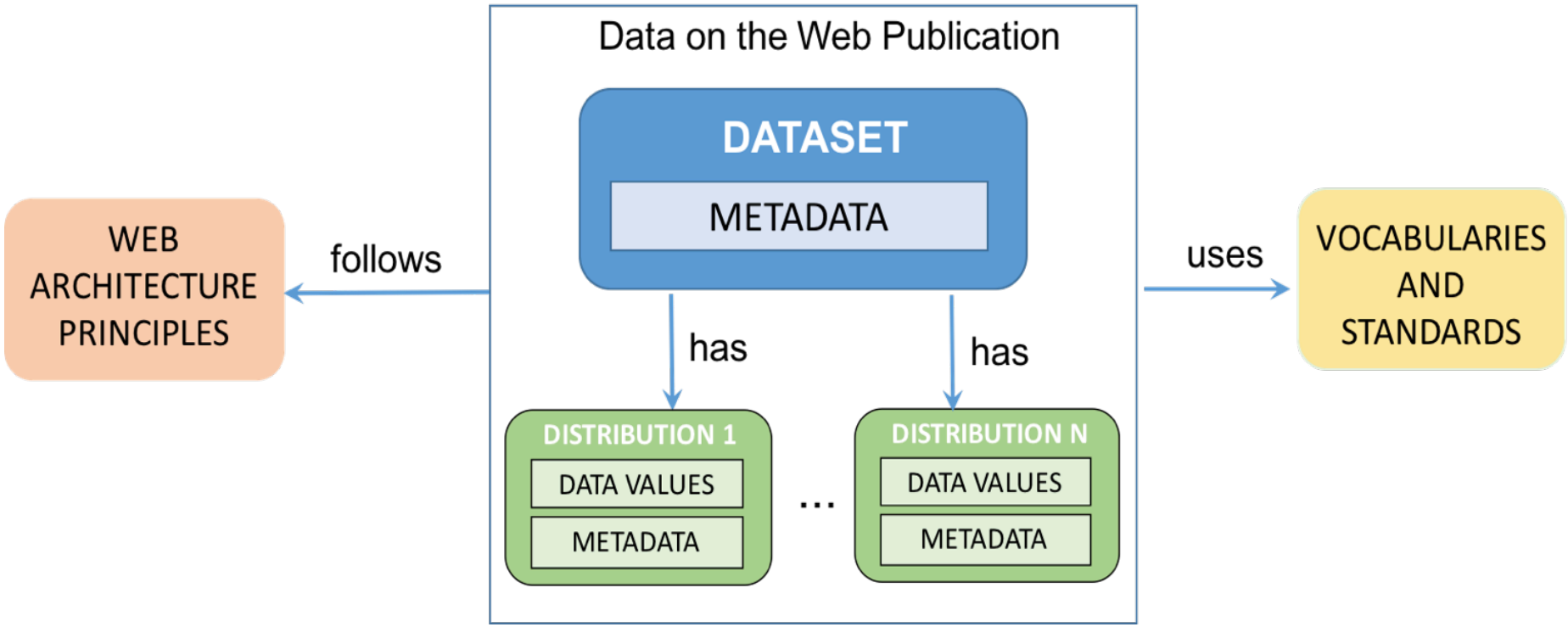
Fuente: http://5stardata.info/

The Linking Open Data cloud diagram



Fuente: http://lod-cloud.net/

Contexto de Datos en la Web

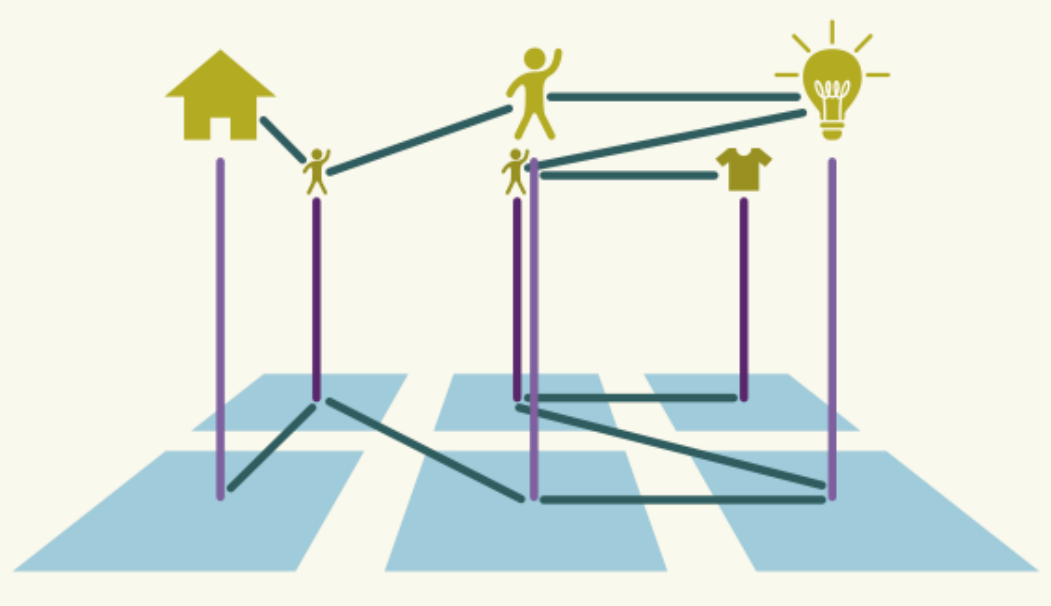


Actores del ecosistema de Datos en la Web

Publicadores de datos: publican y comparten datos

Consumidores de datos: reutilizan los datos y pueden generar nuevos datos

Diversos tipos de recursos de datos (sistemas transaccionales, sensores, dispositivos móviles, documentos…)



Source: http://ceweb.br/livros/dados-abertos-conectados/capitulo-1/

Como posibilitar el reúso de los datos?

Como posibilitar el reúso de los datos?

Un entendimiento común entre los publicadores y consumidores de datos es fundamental.
Sin ese entendimiento el esfuerzo de los publicadores puede ser incompatible con el deseo de los consumidores.

Boas Práticas

Consume datos

Publica datos

Grupo de Trabajo de las Buenas Prácticas para Datos en la Web

La **Misión** del Grupo de Trabajo de las Buenas Prácticas para Datos en la Web, parte de Data Activity de W3C, para:

1. desarrollar el **ecosistema de datos abiertos**, facilitando la comunicación entre publicadores y consumidores de datos;
2. fortalecer **orientación a los publicadores**, que mejorará la consistencia en gestión de datos y promoverá el reúso.
3. **fomentar la confianza de los consumidores** sobre los datos publicados, independiente de la tecnología utilizada, aumentando el potencial para la innovación.

Fonte: https://www.w3.org/2013/dwbp/wiki/Main_Page:

DWBP: Casos de Uso

Data on the Web Best Practices Use Cases & Requirements

W3C Working Group Note 24 February 2015

This version:
http://www.w3.org/TR/2015/NOTE-dwbp-ucr-20150224/
Latest published version:
http://www.w3.org/TR/dwbp-ucr/
Latest editor's draft:
http://w3c.github.io/dwbp/usecasesv1.html
Previous version:
http://www.w3.org/TR/2014/WD-dwbp-ucr-20141014/
Editors:
Deirdre Lee, Derilinx (formerly at Insight@NUIG, Ireland)
Bernadette Farias Lóscio, Centro de Informática - Universidade Federal de Pernambuco, Brazil
Phil Archer, W3C/ERCIM

https://www.w3.org/TR/dwbp-ucr/

DWBP: Casos de Uso

Table of Contents

Publicar datos en la Web es más que apenas "publicar datos"!

Desafíos de Datos en la Web

- Metadatos *(para humanos y máquinas)*
- Licencias de Datos – uso de los datos *(como permitir o restringir acceso a los datos?)*
- Proveniencia & Cualidad de los datos *(cómo añadir confianza a los datos?)*
- Versionamiento de los datos (*acompañar las diferentes versiones)*
- Identificación de los Datos *(identificar datasets y distribuciones)*
- Formatos de los Dados *(cuales formatos de datos utilizar?)*

Desafíos de Datos en la Web

- Vocabularios de Datos *(como promover la interoperabilidad?)*
- Acceso a los Datos *(opciones de acceso a los datos)*
- Preservación de los Datos *(como preservar los datos?)*
- Feedback *(como facilitar la comunicación con usuarios?)*
- Enriquecimiento de los Datos *(añadir valor a los datos)*
- Republicación de los Datos *(reutilizar datos con responsabilidad)*

12 desafíos y 42 requisitos

Data on the Web Best Practices

W3C Recommendation 31 January 2017

This version:
https://www.w3.org/TR/2017/REC-dwbp-20170131/

Latest published version:
https://www.w3.org/TR/dwbp/

Latest editor's draft:
http://w3c.github.io/dwbp/bp.html

Implementation report:
http://w3c.github.io/dwbp/dwbp-implementation-report.html

Previous version:
https://www.w3.org/TR/2016/PR-dwbp-20161215/

Editors:
Bernadette Farias Lóscio, CIn - UFPE, Brazil
Caroline Burle, NIC.br, Brazil
Newton Calegari, NIC.br, Brazil

Contributors:
Annette Greiner
Antoine Isaac
Carlos Iglesias
Carlos Laufer
Christophe Guéret
Deirdre Lee
Doug Schepers
Eric G. Stephan
Eric Kauz
Ghislain A. Atemezing
Hadley Beeman

https://www.w3.org/TR/dwbp/

Audiencia:
Las Buenas Prácticas fueran creadas para atender las necesidades de profesionales que trabajan con gestión de informaciones, desarrolladores, y grupos como investigadores interesados en compartir y reutilizar datos en la Web

Best Practice 1: Provide metadata

Provide metadata for both human users and computer applications.

Why

Providing metadata is a fundamental requirement when publishing data on the Web because data publishers and data consumers may be unknown to each other. Then, it is essential to provide information that helps human users and computer applications to understand the data as well as other important aspects that describes a dataset or a distribution.

Intended Outcome

Humans will be able to understand the metadata and computer applications, notably user agents, will be able to process it.

Possible Approach to Implementation

Possible approaches to provide *human-readable metadata:*

- to provide metadata as part of an HTML Web page
- to provide metadata as a separate text file

Possible approaches to provide *machine-readable metadata:*

- machine-readable metadata may be provided in a serialization format such as Turtle and JSON, or it can be embedded in the HTML page using [HTML-RDFA] or [JSON-LD]. If multiple formats are published separately, they should be served from the same URL using content negotiation and made available under separate URIs, distinguished by filename extension. Maintenance of multiple formats is best achieved by generating each available format on the fly based on a single source of the metadata.
- when defining machine-readable metadata, reusing existing standard terms and popular vocabularies are strongly recommended. For example, Dublin Core Metadata (DCMI) terms [DCTERMS] and Data Catalog Vocabulary [VOCAB-DCAT] can be used to provide descriptive metadata. Such vocabularies are designed to be very flexible so it is often helpful to use a specific *profile* of a vocabulary such as the European Commission's DCAT-AP

EXAMPLE 1

Human-readable

Example page with a human-readable description of an available dataset.

Machine-readable

Example file with a machine-readable description of an available dataset.

How to Test

Check if human-readable metadata is available.

Check if the metadata is available in a valid machine-readable format and without syntax error.

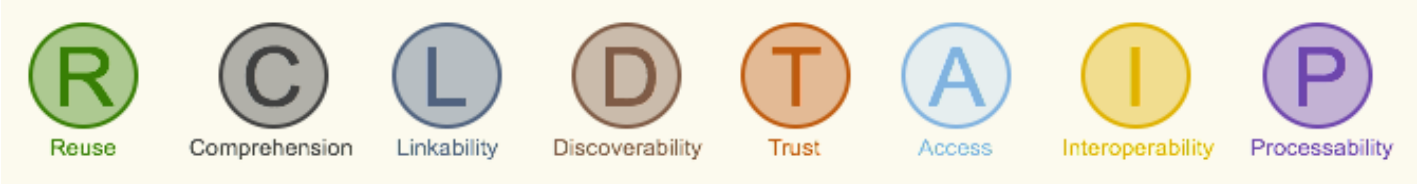
Evidence

Relevant requirements: R-MetadataAvailable, R-MetadataDocum, R-MetadataMachineRead

Benefits

Beneficios DWBP

Cada beneficio representa una mejora en como conjuntos de datos (datasets) son disponibles en la Web



Reuse

BP: Provide data license information
BP: Provide versioning information
BP: Provide version history
BP: Use non-proprietary data formats
BP: Provide data in multiple formats
BP: Use a trusted serialization format for preserved data dumps
BP: Enrich data by generating new metadata
BP: Provide data provenance information
BP: Provide data quality information
BP: Use persistent URIs as identifiers

Discoverability

BP: Provide descriptive metadata
BP: Use persistent URIs as identifiers
BP: Assign URIs to dataset versions and series

Trustworthy

BP: Assess dataset coverage
BP: Assign URIs to dataset versions and series
BP: Provide data up to date
BP: Update the status of identifiers
BP: Gather feedback from data consumers
BP: Provide information about feedback
BP: Provide data provenance information
BP: Provide data quality information

Linkability

BP: Use persistent URIs as identifiers
BP: Assign URIs to dataset versions and series

Processibility

BP: Use machine-readable standardized data formats
BP: Enrich data by generating new metadata

Comprehension

BP: Provide metadata
BP: Provide locale parameters metadata
BP: Provide structural metadata
BP: Provide descriptive metadata

Accessibility

BP: Provide bulk download
BP: Follow REST principles when designing APIs
BP: Provide real-time access
BP: Maintain separate versions for a data API
BP: Assess dataset coverage

Interoperability

BP: Use standardized terms
BP: Re-use vocabularies

Best Practice 1: Provide metadata

Metadata must be provided for both human users and computer applications

Why

Providing metadata is a fundamental re
lishers and data consumers may be unk
that helps human users and computer a
aspects that describes a dataset or a di

Intended Outcome

Human-readable metadata will enable h
metadata will enable computer applicati

Possible Approach to Implementatio

Possible approaches to provide *human*

- to provide metadata as part of an H
- to provide metadata as a separate

Possible approaches to provide *machin*

- machine readable metadata may b
  it can be embedded in the HTML pa
  published separately, they should b
  nance of multiple formats is best ac
  a single source of the metadata.
- when defining machine readable m
  ies are strongly recommended. For example, Dublin Core Metadata (DCMI) terms [DC-TERMS]
  and Data Catalog Vocabulary [VOCAB-DCAT] should be used to provide descriptive metadata.

Beneficios de las Buenas Prácticas

- **Comprensión**: Las personas son capaces de entender mejor la estructura de los datos, el significado de los datos, así cómo los metadatos y la naturaleza del conjunto de datos.
- **Procesabilidad**: Las máquinas serán capaces de procesar automáticamente y manipular los datos de un conjunto de datos.
- **Descubrimiento:** Las máquinas automáticamente serán capaces de descubrir un conjunto de datos o los datos en los contenidos.
- **Reúso**: las oportunidades de reúso de los datos o conjuntos de datos por diferentes grupos irán aumentando.

Best Practice 10: Use persistent URIs as identifiers of datasets

ReSpec

Datasets must be identified by a persistent URI.

Why

Adopting a common identification syster
by any stakeholder in a reliable way. The
and reuse.

Developers may build URIs into their co
dereference to the same resource over t

Intended Outcome

Datasets or information about datasets v
status, availability or format of the data.

Possible Approach to Implementatior

To be persistent, URIs must be designec
creating a Web site designed for human
topic, see, for example, the European C
to many other resources.

Where a data publisher is unable or unw
native approach is to use a redirection s
These provide persistent URIs that can
ephemeral. The software behind such services is freely available so that it can be installed and managed locally if required.

Digital Object Identifiers (DOIs) offer a similar alternative. These identifiers are defined independently of any Web technology but can be appended to a 'URI stub.' DOIs are an important part of the digital infrastructure for research data and and libraries.

Beneficios de las Buenas Prácticas

- **Vinculación de links**: será posible crear links entre recursos de datos (conjuntos de datos e ítems de datos).
- **Interoperabilidad**: será más fácil obtener consenso entre publicadores y consumidores de datos.
- **Confianza:** mejorará la confianza quelos consumidores tienen en relación a los conjuntos de datos.
- **Acceso:** personas y máquinas serán capaces de tener acceso a datos actuales y en diferentes formatos.

Cómo participar ahora?

TABLE OF CONTENTS

DWBP Implementation Report

W3C Document 29 January 2017

Editors:
Bernadette Farias Lóscio, CIn - UFPE, Brazil
Caroline Burle, NIC.br, Brazil
Newton Calegari, NIC.br, Brazil

Abstract

This document reports on evidence and implementations of the Data on the Web Best Practices Candidate Recommendation. In particular, it demonstrates that the DWBP are already in use and are also implementable.

Status of This Document

This document is merely a W3C-internal document. It has no official standing of any kind and does not represent consensus of the W3C Membership.

1. Introduction

One of the main goals of the Data on the Web Best Practices (DWBP) is to facilitate interaction between publishers and consumers of data on the Web. A set of 35 Best Practices were created to cover different challenges related to data publishing and consumption, such as Metadata, Data licenses, Data provenance, Data quality, Data versioning, Data identification, Data formats, Data vocabularies, Data access and APIs, Data preservation, Feedback, Data enrichment and Data republication.

Fuente: http://w3c.github.io/dwbp/dwbp-implementation-report.html

Nuevo Grupo de Data Activity de W3C

W3C

Dataset Exchange Working Group Charter

The **mission** of the Dataset Exchange WG is to:

- Revise the Data Catalog Vocabulary, DCAT, taking account of related vocabularies and the extensive work done in developing a number of its application profiles.
- Define and publish guidance on the use of application profiles when requesting and serving data on the Web.

Join the Dataset Exchange Working Group.

Start date	04 May 2017
End date	30 June 2019
Chairs	Caroline Burle, NIC.br, Karen Coyle, Dublin Core Metadata Initiative
Team Contacts	Phil Archer (0.2 FTE), supported by the VRE4EIC project
Meeting Schedule	**Teleconferences:** 1-hour calls will be held weekly **Face-to-face:** twice per year, expected to include the W3C's annual Technical Plenary week.

2. Scope

DCAT is formulated as an RDF vocabulary and is expected to remain so, however, the working Group is agnostic about data formats. Methods for expressing DCAT in other (existing) formats are in scope.

Government data, scientific research data, industry/enterprise and cultural heritage data, whether shared openly or not, are all explicitly in scope.

2.1 Input Documents

The following documents *SHOULD* be considered by the Working Group as direct inputs to the specifications to be developed.

W3C documents

- DCAT and the HCLS Community profile
- The Data Quality and Dataset Usage vocabularies
- The Smart Data & Smarter Descriptions (SDSVoc) workshop report, in particular the section on content negotiation by application profile.
- Data on the Web Best Practices

https://www.w3.org/2017/dxwg

Materiales de Referencia

Fuente: http://ceweb.br/publicacoes/indice/

Curso de Buenas Prácticas de Datos en la Web

Gracias!

www.ceweb.br - www.cin.ufpe.br

 cburle@nic.br

 bfl@cin.ufpe.br

 newton@nic.br

 @carolburle

 @bernafarias

 @newtoncalegari

El caso de Costa Rica – Justicia Abierta

Carlos Morales Castro

Proceso inicial

Educar sobre el tema de datos abiertos

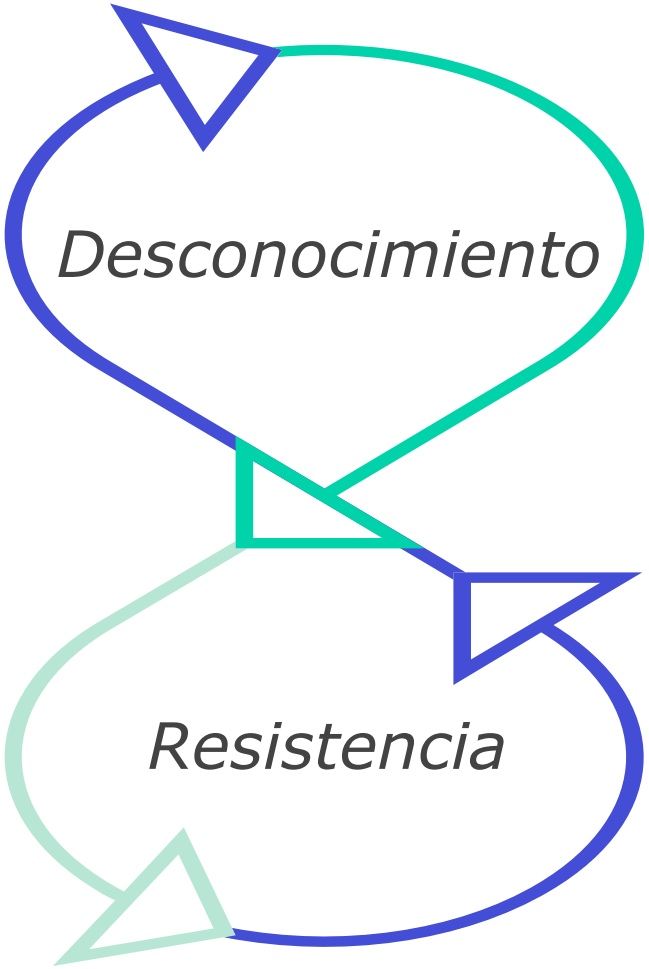
01 *Análisis de lo que tenemos*

02 *Consulta pública*

03 *Recolección, limpieza y verificación de datos*

04 *Publicación*

Dificultades

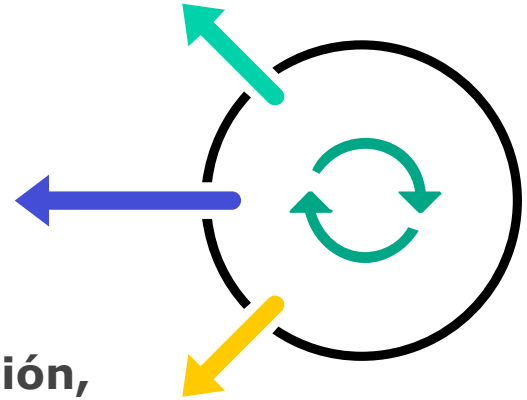


Proceso actual (Informal)

Solicitud del conjunto de datos

Análisis de fuentes de datos

Recolección, limpieza, verificación y publicación de datos

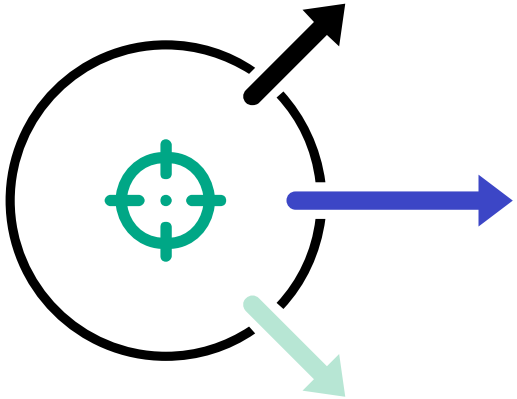


Dificultades

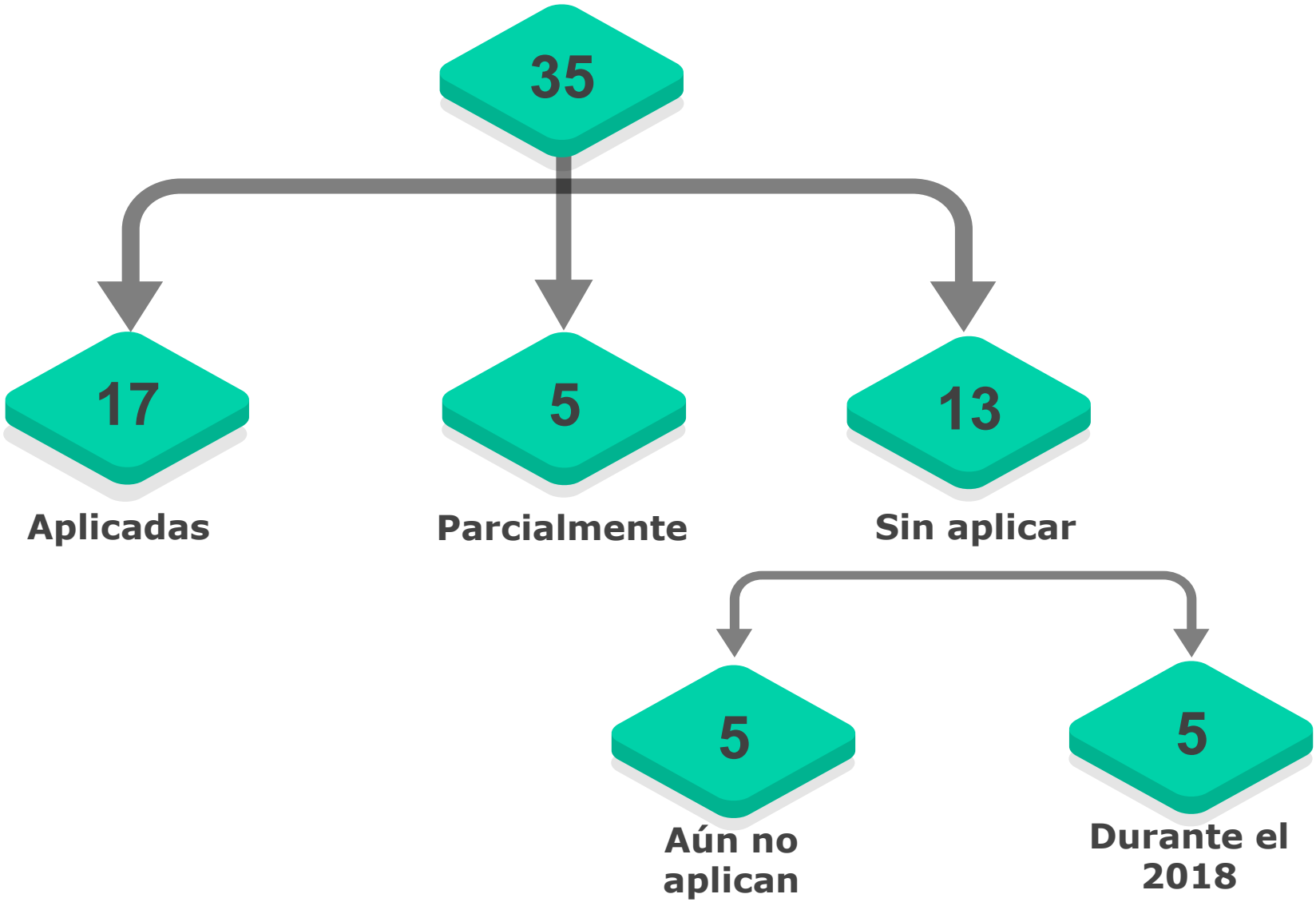
Procesamiento de datos no estructurados

Procesos para la creación de datos

Normativa



Buenas prácticas



Pasos futuros

01

Incorporar el API

02

Educar sobre la importancia de mejorar los procesos de creación de datos

03

Agregar más conjuntos de datos

Gracias!

http://www.poder-judicial.go.cr/justiciaabierta/index.php/datos-abiertos

cmoralesc@poder-judicial.go.cr