



The IPTC Photo Metadata (PMD) and data mining

4 May 2023



David Riecks



Michael Steidl

Co-Leads of
IPTC Photo Metadata Working Group

Digital News, Enabled.

The IPTC is the global standards body of the news media. We provide the technical foundation for the news ecosystem.

OUR STANDARDS

FOR DEVELOPERS



Photo Metadata



Video Metadata Hub



Media Topics



RightsML

About 60
members

7 working
groups on
standards



The Metadata Sea and its Photo Metadata Bay



The Islands in the Photo Metadata Bay

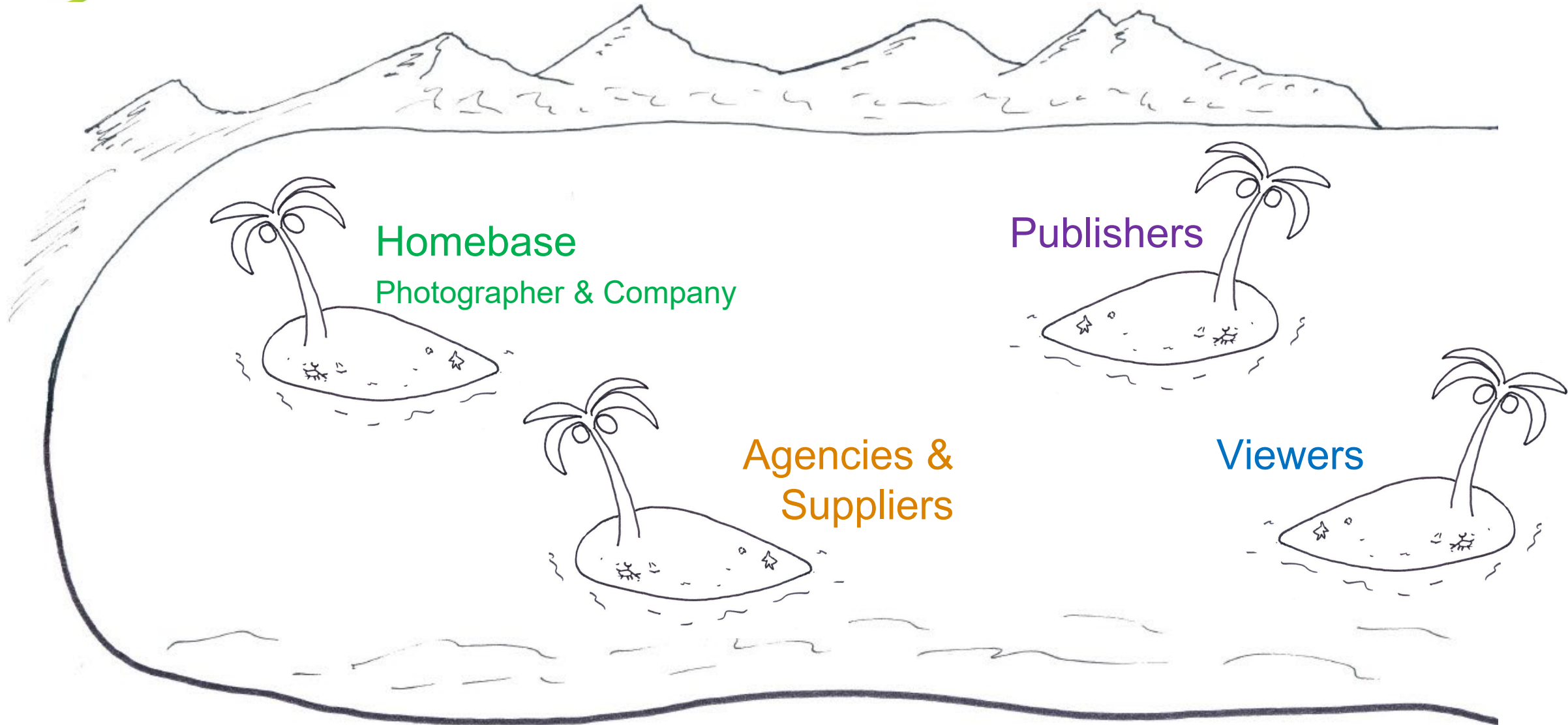
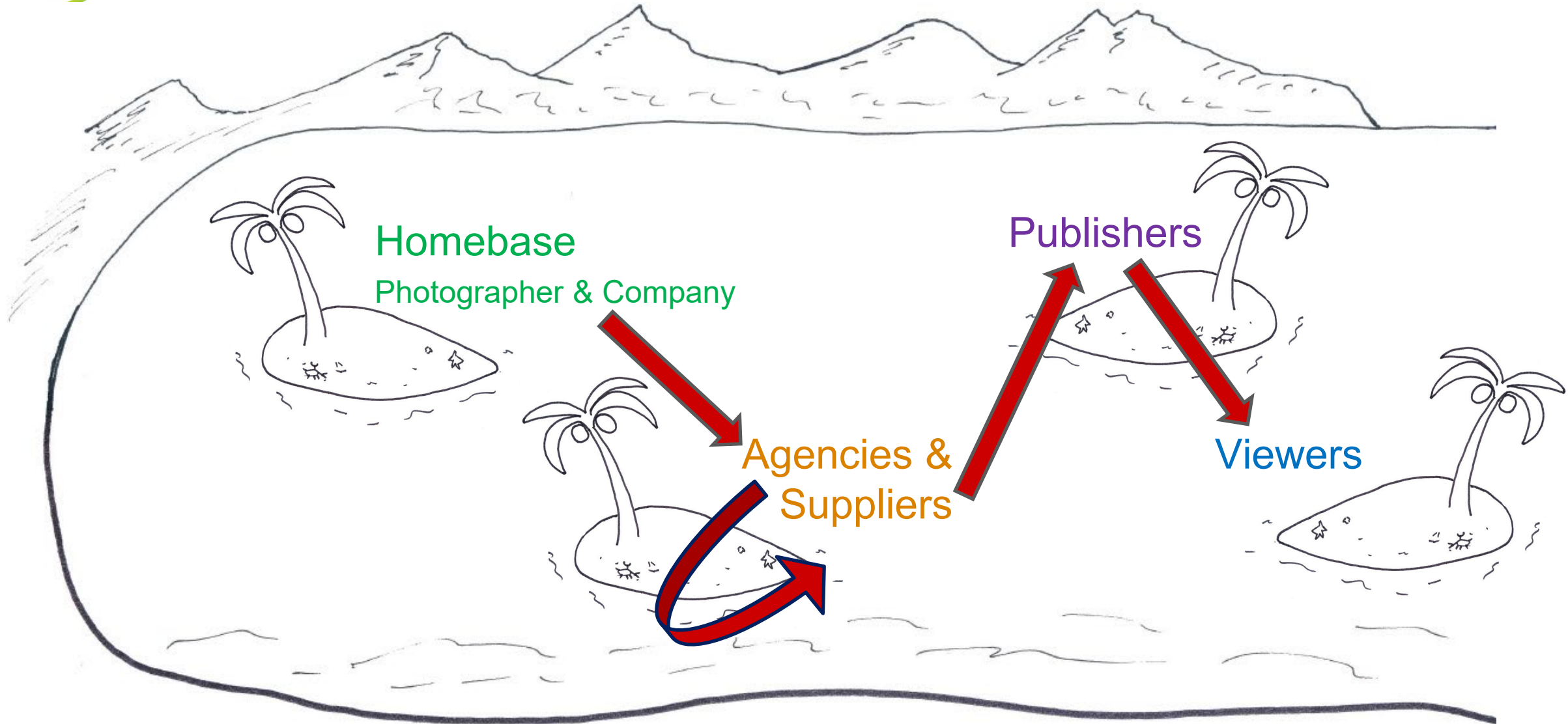


Photo (Metadata) flow



Goal of a PMD flow across islands

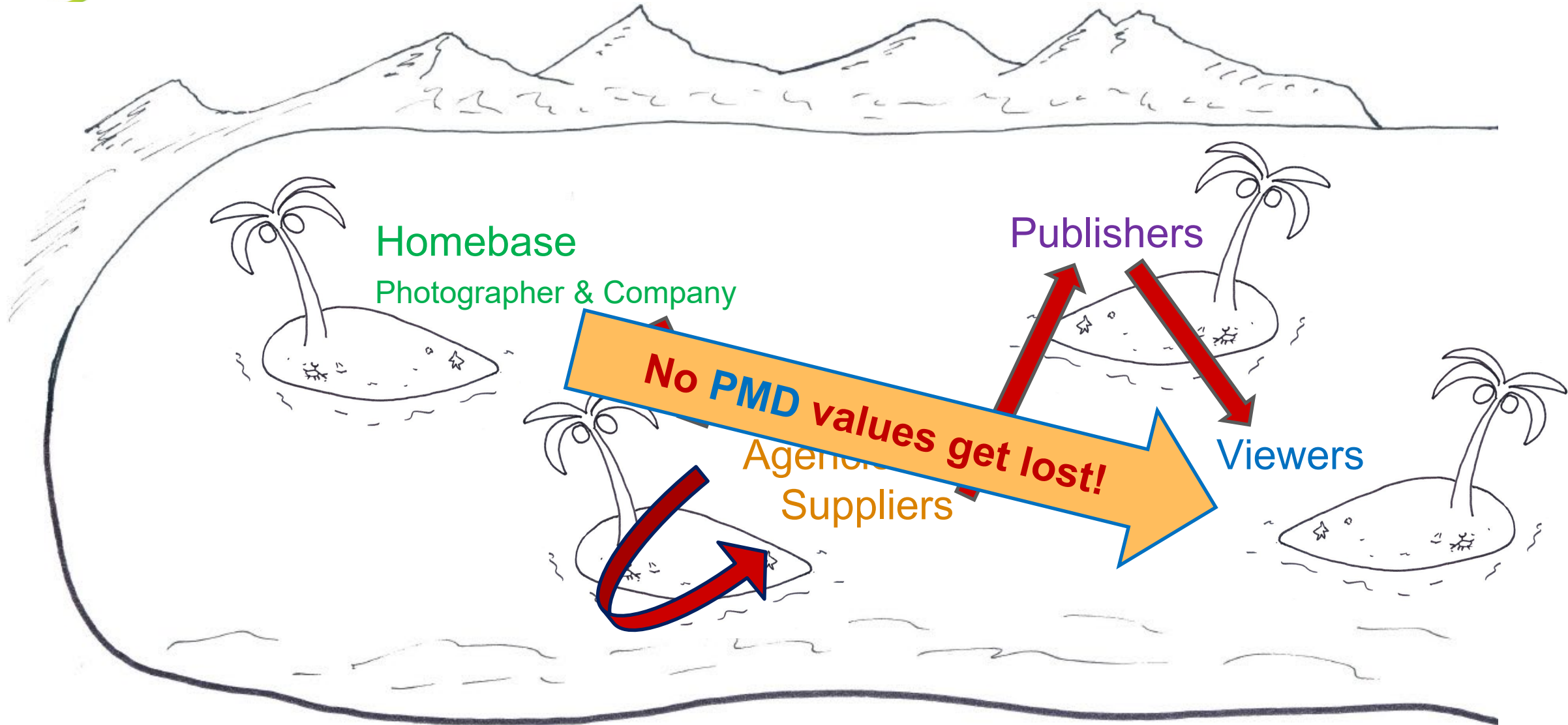




Photo Metadata Standards

Characteristics and goals of the [IPTC Photo Metadata Standard](#):

- Bound to the file of the image the meta-assertions are about = embedded
- Supporting backward compatibility: no PMD property was deleted, only some are earmarked as “deprecated”.
- Properties are created and maintained based on the needs of photographers and photo businesses (photo agencies, stock photo companies, ...).

Not only IPTC lives in the Photo Metadata Bay:

- [Exif](#) focuses on metadata created by cameras, a few are [mapped](#) to IPTC PMD
- [PLUS](#) specifies rights and licensing related metadata, IPTC adopted a bunch
- [Schema.org](#) specifies metadata for web content, some are [mapped](#) from IPTC
- [CAI](#) and [C2PA](#) make existing metadata secure – and added properties



IPTC & PLUS thinking about data mining (1)

- Data mining can be understood as
 - search engines investigate web content to display search results: this use is in the range of accepted to liked.
 - analyzing web content to find patterns, trends, characteristics of the data: this is ok for scientific research
 - analyzing web content to build commercial business on the found patterns, trends, characteristics of the data: it should be possible to prohibit this use.



IPTC & PLUS thinking about data mining (2)

A metadata property about data mining should ...

- ... have a small set of values – a controlled vocabulary – else users won't use it.
- ... have values being readable for humans (easy to understand labels) and for machines (globally unique identifiers of values) - as assets are collected by web crawlers.
- ... have values for a clear Allowed and Prohibited of data mining^(shared semantics?)
- ... have a value for excluding AI generative training from Allowed (and maybe others for excluding more ...)
- ... have a system for using refined machine-readable assertions, using another property with a focus on this kind of use.



IPTC & PLUS data mining property DRAFT

- A single property (0..1) with a value (identifier) from this vocabulary:
 - **Allowed** (TDMRep: tdm-reservation=0)
 - **Allowed, except for the purpose of AI/ML training** (c2pa.ai_training->use="notAllowed"; TDMRep: tdm-reservation=1 must be used with a TDMRep policy)
 - **Allowed, except for the purpose of AI/ML generative training** (c2pa.ai_generative_training->use="notAllowed"; TDMRep: tdm-reservation=1 must be used with a TDMRep policy)
 - **Prohibited** (TDMRep: tdm-reservation=1)
 - **Allowed with constraints, read the human-readable PLUS *Other Constraints* property**
 - **Allowed with constraints, read the machine-readable IPTC *Embedded Encoded Rights Expression* property** (TDMRep: tdm-reservation=1 + TDMRep policy)
 - **Allowed with constraints, resolve and read the machine-readable IPTC *Linked Encoded Rights Expression* property** (TDMRep: tdm-reservation=1 + TDMRep policy)
 - **Unspecified** = Nothing about data mining is specified for this asset



This property will be shown at <https://getpmd.iptc.org>



Get IPTC Photo Metadata (2022.1)

Option A: ... embedded into an image on the web

Enter the URL of the image file:

(If this field is left empty the metadata of the IPTC reference image of the Photo Metadata Standard 2022.1 is shown. This image file comprises all metadata fields defined by that standard.)

Select how the IPTC Photo Metadata fields are shown:

- As single list**, sorted by names
- Grouped by common topics** (General content, persons, locations, ..., rights, licence, administrative issues)
- Selected fields **relevant for photos shown as result by search engines**
- Grouped by the technical formats** used in the image files (**IIM, XMP and Exif**)

An IPTC field is shown ... only if it **has a value** if it **has a value or no value**

Shown field names: **IPTC Photo Metadata Standard** names Field IDs of **XMP, IIM, Exif** - only with "Grouped by the technical formats ..."

Get Photo Metadata of Web Image