

La parte no informática

1. ¿Que es calidad en los datos?
 - a. Que los datos nos sirvan para lo que queremos hacer: "fitness for use"
 - La perfección no existe.
 - El dato malo tampoco; un dato no es bueno o malo intrínsecamente, sino que su bondad depende del uso.
 - Definir nuestros objetivos en términos de uso de los datos (actuales y potenciales) es un elemento clave a la hora de que prestar atención y con que prioridad
 - Implicaciones como proveedores de datos
 - Implicaciones como "consumidores" de datos
2. Como saber (y hacer saber) que un datos es "bueno"
 - a. El problema es saber si un datos es bueno (en general este es un problema asociado a todo lo que es información)
3. Más contexto

Conocimiento explícito e implícito

- a.
 - El conocimiento –contexto si se prefiere– es algo muy difícil de aprehender pero vital para un uso óptimo de los ejemplares y las colecciones.
 - La documentación proporciona contexto.
 - Las tecnologías informáticas están devolviendo las colecciones al la escena de la ciencia actual y de las preocupaciones sociales. Sin embargo, este procesado de conlleva una descontextualización de los datos.
 - Esta situación hace que una buena documentación sea más importante que nunca.

Hay que luchar contra el síndrome de "todo el mundo sabe eso" (y no se documenta)

Aspectos generales



- b. Cuando hablamos de "documentar el dato" estamos hablando de metadatos
- c. Ejemplos de metadatos
 - i. De bases de datos en línea
 - 1) <http://www.gbif.es/datos/colecciones>
 - ii. Foto
<file:///I:/AFOTOS/Taller%20Imágenes%202009>
 - iii. Registro de colecciones y proyectos de biodiversidad en España
http://www.gbif.es/ic_colecciones.php?ID_Coleccion=9575
<http://es.mirror.gbif.org/datasets/resource/239/>

-	En definitiva, proporcionar metadatos es dar al usuario los elementos necesarios para establecer si un datos le sirve o no.
---	---

4. ¿Cual es la manera de que los datos que se introduzcan sean de buena calidad?
 - a. Ideas
 - i. Disponibilidad de herramientas
 - b. Que los investigadores introduzcan los datos para su propio beneficio
 - i. Etimatic> HZL
<http://www.gbif.es/hzl/hzl.php>
 - c. Fundamental que el primer beneficiario de la digitalización sea el propio digitalizador

5. ¿Más tecnología implica más calidad?
 - a. Vocabularios controlados
 - b. Interfaces avanzados
 - c. -- sin apoyo; sin incentivos, la tecnología no funciona
 - d. Acceso a datos - como de fácil es sacar los datos de un sistema es el primer criterio de elección de un software
 - e. "Bus test"
6. Que queremos y que es lo que más cuesta
 - a. Hacer lo que hacemos, pero mejor
 - b. Que es lo que más cuesta
 - i. En disponibilidad
 - 1) Buena identificación
 - ii. En recursos (tiempo)
 - 1) Manipular el material (problema específico de la informatización de las colecciones)
 - c. Que te paguen por digitalizar... o hacer lo que hacemos, digitalizando con la idea de reutilizar la información
 - i. Etiquetas, préstamos, informes,...

7. Prevenir y curar

El contexto de la digitalización de colecciones, tener claras las prioridades:

- Conservar
- Acceder
- Informatizar

Mantener la información original ...

..y distinguible de la interpretada

Prevenir errores

Consistencia -> guías, manuales, vocabularios controlados

Ejemplo:

ISO 3166

http://en.wikipedia.org/wiki/ISO_3166-1_alpha-2

Publicaciones (botánicas)

<http://www.ipni.org/ipni/publicationsearchpage.do>

Corrección temprana de errores

La corrección es más sencilla y eficaz cuanto más cercanos están los procesos de control y corrección a los de introducción de datos

8. Optimización: el método esloveno

- a. Las claves
 - i. El primer paso de la digitalización es fotografiar el ejemplar
 - ii. El proceso se descompone en pasos secuenciales de complejidad mínima
 - iii. Cada paso del proceso hace el control de calidad del anterior
- b. Valores añadidos
 - i. Mas eficiente
 - ii. Desacopla el trabajo en la colección del proceso de digitalización
- c. PPT



i.

Kotarec

ii. <http://enbi.maich.gr/pastworksh/pdf2/Kotarec.pdf>

9. Guía de digitalización de imágenes

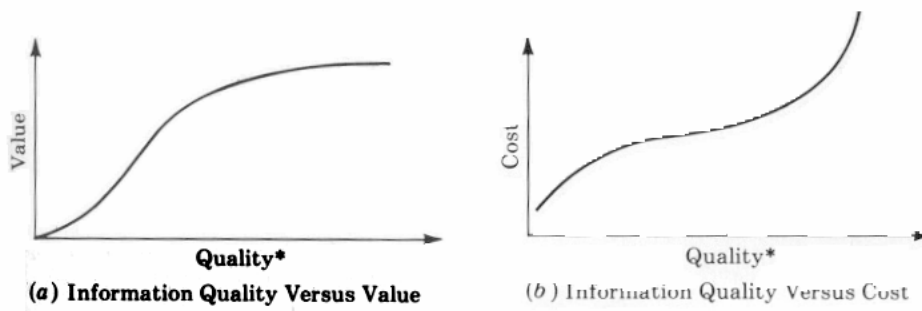
A Manual of Best Practice: digital Imaging of Biological Type Specimens

<http://wiki.gbif.org/nodeswiki/wikka.php?wakka=NodesGuide>

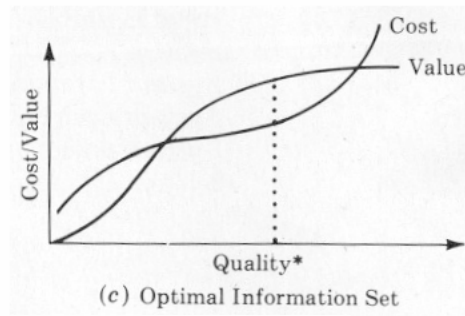
Taller sobre imágenes digitales para estudios de biodiversidad

<http://www.gbif.es/formaciondetalles.php?IDForm=58>

10. Aparte: calidad y coste



a.



*Information quality = $F(\text{detail, age, accuracy, relevance})$