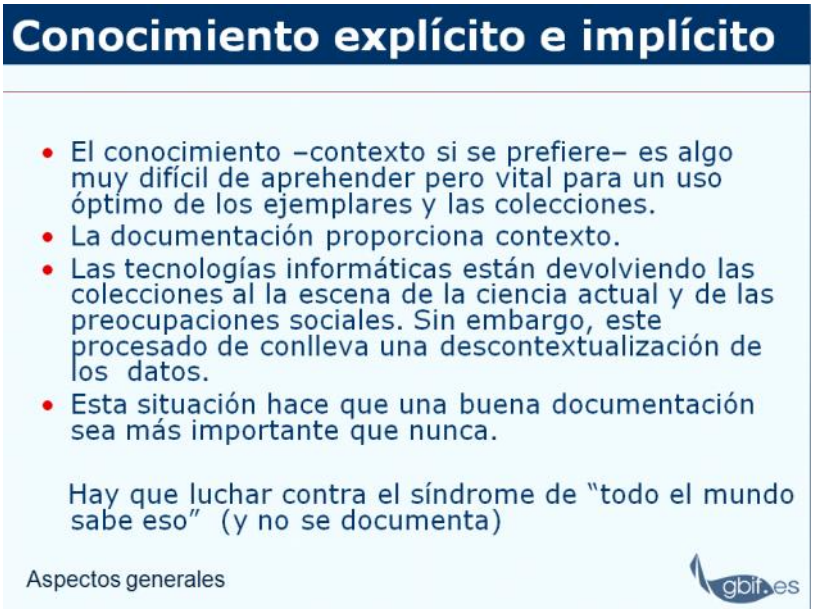
Hacia la calidad de datos: la parte no informática

martes, 25 de noviembre de 2008

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
2. Más contexto
    a. 

    b. visión y misión
        i. Donde queremos estar
        ii. Que tenemos que hacer para estar donde queremos

        "vision without mission is a daydream.. a mission without vision is a nightmare"

        Definir que 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
3. Que el dato esté, que sea "bueno"
    a. El problema es saber si un datos es bueno
        i. Importancia de los "metadatos" (=datos sobre datos)
            1) Ejemplo
                a) El nombre de la persona que hace una identificación es metadatos del nombre de la identificación
            2) Ejemplo
                a) La dirección postal de una colección es metadato de la colección
            3) Ejemplo

                ISO 3166 es metadato de ESP para nombre de país

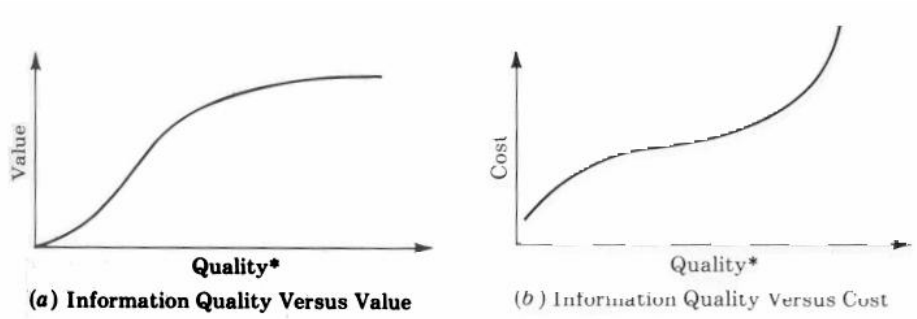
    En otras palabras, proporcionar metadatos es dar al ususario los elementos necesarios para establecer si un datos le sirve o no.
4. ¿Cual es la manera de que los datos que se introduzcan sea de buena calidad?

a. Ideas
    i. Disponibilidad de herramientas
b. Que los investigadores introduzcan los datos para su propio beneficio
    i. Etimatic …….> Herbar ligero
        http://www.gbif.es/herbar/Herbar_ligero.php
c. Fundamental que el primer beneficiario de la digitalización sea el propio digitalizador

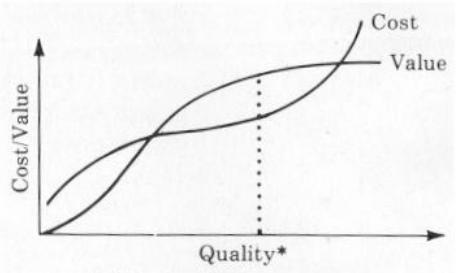
5. ¿Más tecnología implica más calidad?
    a. Vocabularios controlados
    b. Interfaces avanzados
    c. -- sin apoyo; sin incentivos, la tecnología no funciona
    d. Acceso a datos - como de fácil es sacar los datos de un sistema es el primer criterio de elección de un software
    e. Bus test
6. 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
7. Prevenir y curar
    El contexto de la digitalización
        Conservar
        Acceder
        Informatizar
    Prevenir errores
        Consistencia -> guias, manuales, vocabularios controlados
    Mantener la información original y distinguible de la interpretada
    Procesos de control y corrección cercanos al de introducción de datos
    Distinguir la información original de la interpretada
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=38

10. Aparte: calidad y coste

a.



(a) Information Quality Versus Value (b) Information Quality Versus Cost



(c) Optimal Information Set

***Information quality = *F*(detail, age, accuracy, relevance)**