



Importancia de la Usabilidad en los Sistemas de Flujo de Trabajo Científicos

El caso del Laboratorio Virtual de Biotecnología de la UPPue

*Dr. Abraham Nieva de la Hidalga
Dra. María Auxilio Medina Nieto*

Día Virtual sobre el Laboratorio Abierto de
Usabilidad y Experiencia de Usuario (LAbUX)

Noviembre 13 del 2015



Contenido

1. Manejo de datos en la investigación
2. Sistemas de flujos de trabajo científicos
 - 2.1 Beneficios
 - 2.2 Barreras
3. Laboratorio virtual de biotecnología
4. Estudios de usabilidad para el LVB



1. Manejos de datos en la investigación

Algunas de las áreas de interés son:

- Genómica, proteómica, metabolómica
- Secuenciación de genes
- Datos de cambio climático
 - Física a gran escala (astronomía, física de partículas, etc.)



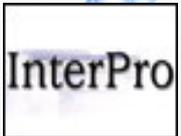


1. Manejos de datos en la investigación



Entre los recursos de información se encuentran:

- NAR: 1552 bases de datos
- Genbank: 172 millones de secuencias
- WSG: 774 billones de pares base



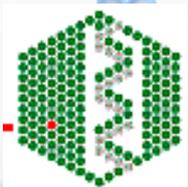


1. Manejos de datos en la investigación



Los recursos se almacenan y distribuyen en medios diferentes como:

- Bases de datos locales
- Páginas web estáticas
- Aplicaciones web dinámicas
- Servidores FTP
- En documentos *.txt, *.pdf, *.doc, *.xls o *.ppt
- Servicios web





2. Sistemas de flujos de trabajo científicos

Una investigadora o investigador realiza actividades propias de su rol como:

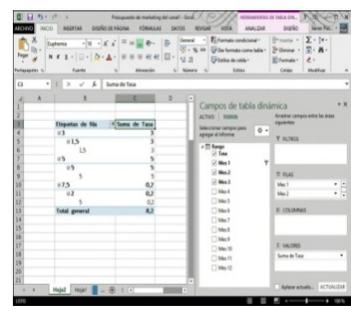
- Emplear y combinar recursos públicos y propios
- Utilizar programas y computadoras para analizar datos
- Dar a conocer los resultados de su trabajo a colaboradores y colegas



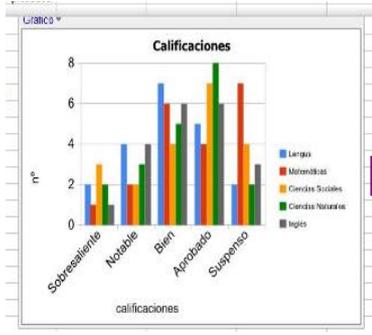
Flujo de trabajo implícito de investigadores



Trabajo en laboratorio



Registro de observaciones

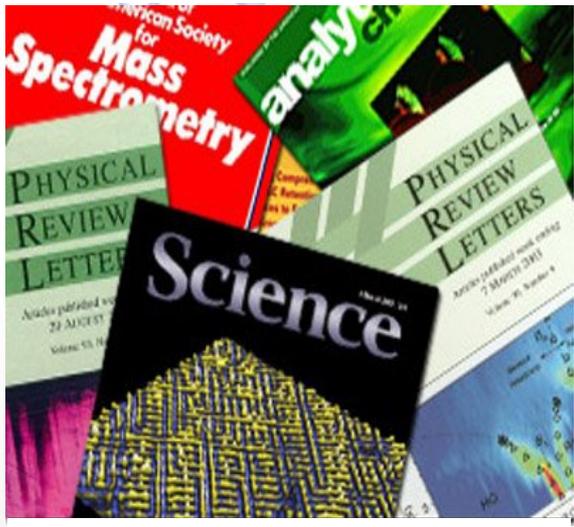


Selección de datos



$$I_M^{(1)} = \prod_{J=0}^{M-1} \prod_{1 \leq i < j \leq 4} \Gamma((pq)^{r_{MJ}} y_i y_j; p, q) \prod_{5 \leq i < j \leq 8} \Gamma((pq)^{r_{MJ}} y_i y_j; p, q) \\ \times \Gamma((pq)^s; p, q)^{N-1} \frac{(p; p)_\infty^N (q; q)_\infty^N}{2^N N!} \int_{T^N} \prod_{1 \leq i < j \leq N} \frac{\Gamma((pq)^s z_i^{\pm 1} z_j^{\pm 1})}{\Gamma(z_i^{\pm 1} z_j^{\pm 1})} \\ \times \prod_{j=1}^N \frac{\prod_{i=1}^4 \Gamma((pq)^{r_{qj}} v^{-2} y_i z_j^{\pm 1}; p, q) \prod_{i=5}^8 \Gamma((pq)^{r_{qj}} v^2 y_i z_j^{\pm 1}; p, q)}{\Gamma(z_j^{\pm 2}; p, q)}$$

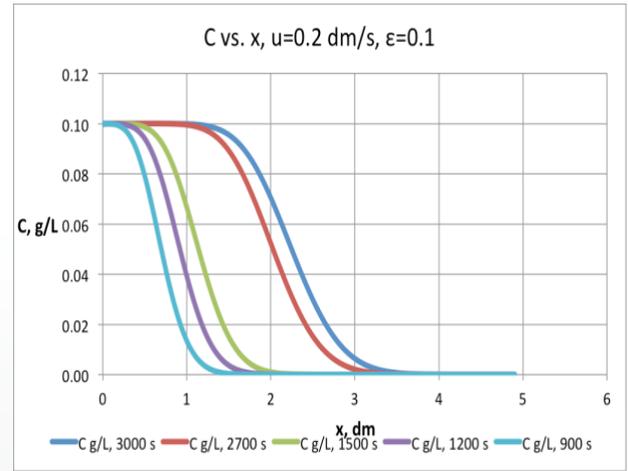
Análisis de datos



Publicación de resultados



Elaboración de documentos de divulgación científica



Creación de modelos



2. Sistemas de flujos de trabajo científicos

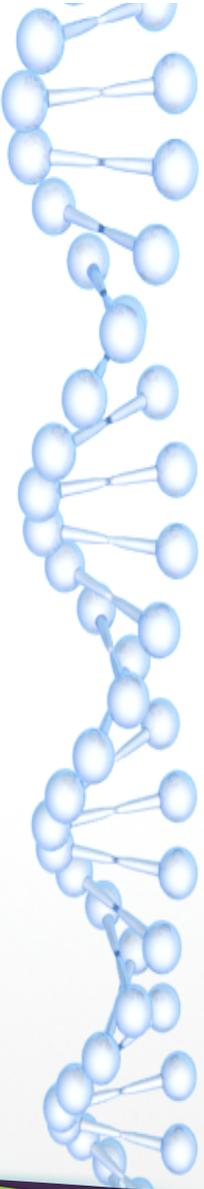
Algunas de las actividades que se realizan en un flujo de trabajo implícito son:

- Uso secuencial de herramientas de análisis
- Manejo de formatos no compatibles de entrada y salida
- Dificultades para registrar, producir y modificar parámetros, procedimientos, herramientas o datos intermedios



2. Sistemas de flujos de trabajo científicos

Un flujo de trabajo implícito es aceptable de manera ocasional, no así en escenarios que requieran análisis de datos a gran escala



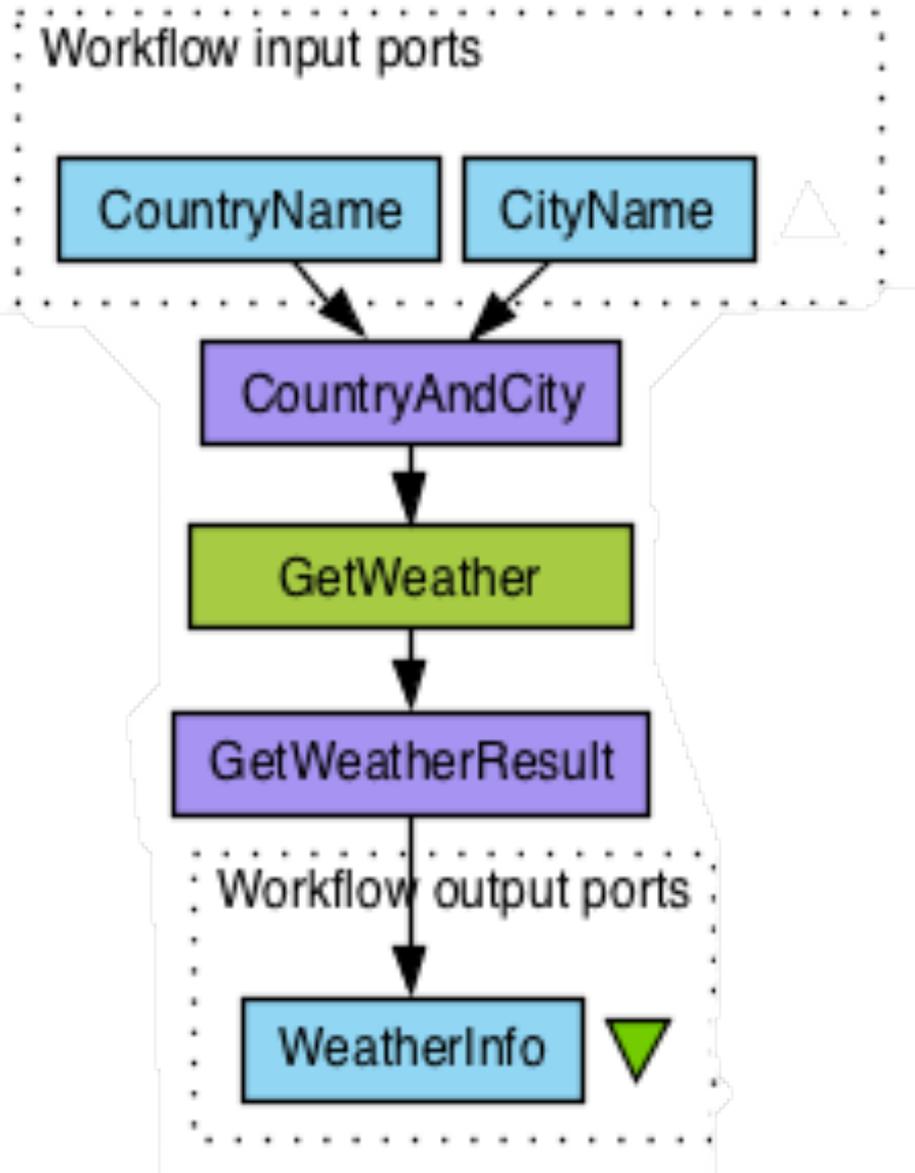


2. Sistemas de flujos de trabajo científicos

Un sistema de flujo de trabajo científico (SFTC) se relaciona con:

- Herramientas sofisticadas para análisis de datos
- Representaciones gráficas del proceso de análisis
- Combinación de servicios para analizar y administrar datos locales y/o remotos
- Iteración en diversas fuentes de datos
- Ejecución automática y repetible

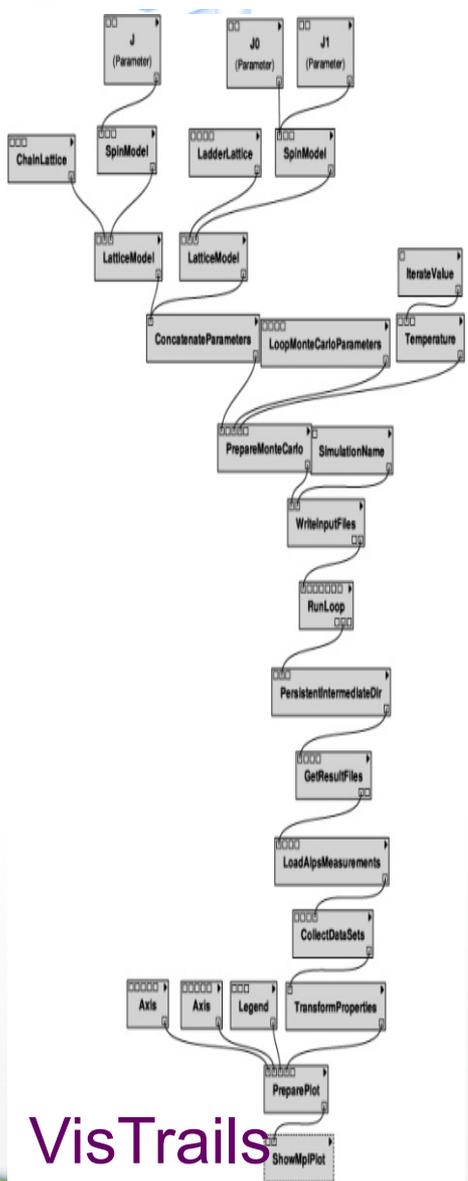
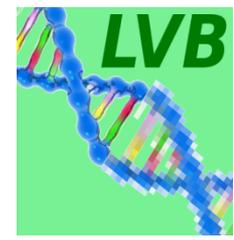
Ejemplo 1: SFTC para obtener el pronóstico del tiempo para una ciudad y país utilizando **taverna**



- La caja verde es un **servicio web**
- Las cajas moradas son **servicios XML locales** para organizar y extraen datos
- Las cajas azules representan **puertos de entrada y salida de datos**

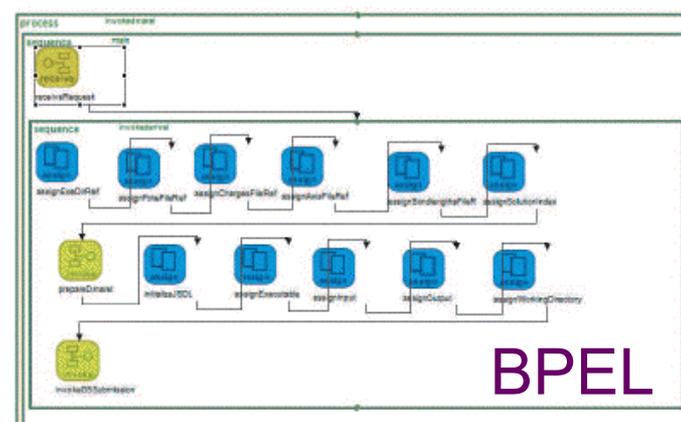
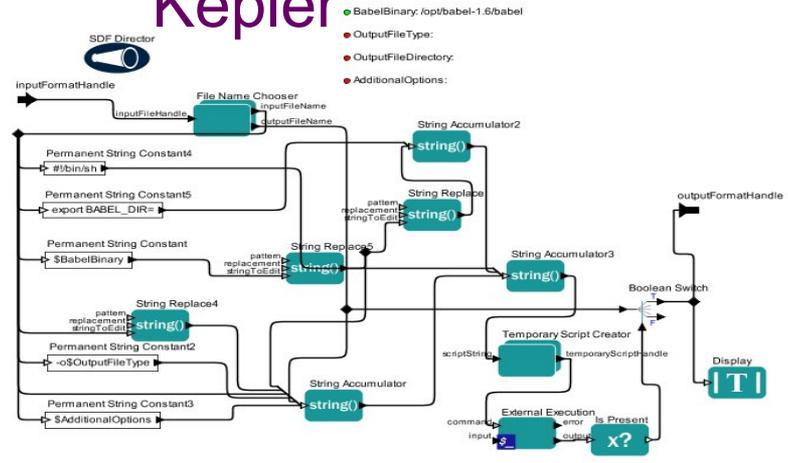


Ejemplos de SFTC

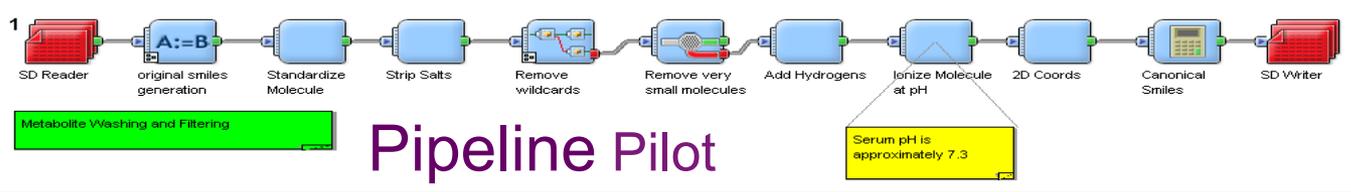


VisTrails

Kepler

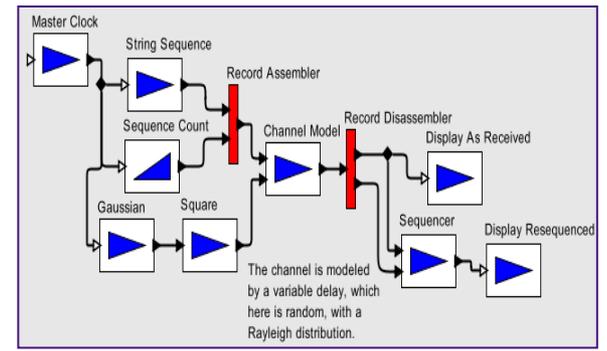
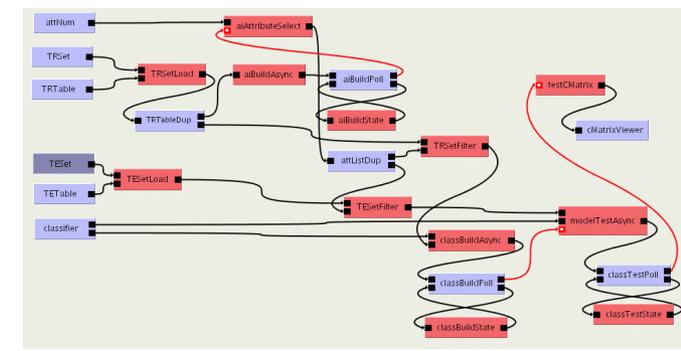


BPEL



Pipeline Pilot

Triana



Ptolemy II



Tipos de usuarios



Avanzados: Diseñan y construyen SFTC

The screenshot shows the Taverna Workbench 2.2.0 interface. On the left is the 'Service panel' with a search filter and a list of 'Available services' including 'Beanshell', 'Nested workflow', 'Rshell', 'Spreadsheetimport', and 'String constant'. Below this is the 'Workflow explorer' showing a tree view of the workflow components. The main area is the 'Workflow diagram' showing a complex flow of services: 'getMolarVolumeAndRefractivity' (input: compound) leads to 'createURL', then 'getMolarVolumeRefractivity', 'parseOutput', and 'getAbrahamDescriptors' (output: molarVolume, molarRefractivity, compound). This is followed by another 'createURL' step, 'getAbrahamDescriptors' (input: molarVolume, molarRefractivity, compound), and finally 'abrahamDescriptors' (output: volume, refractivity, chemspiderid).

my experiment beta

The screenshot shows the myExperiment website interface. The top navigation bar includes 'Users', 'Groups', 'Workflows', 'Files', 'Blogs', and 'Forums'. The main content area displays a workflow entry titled 'Workflow Entry: Microarray CEL file to candidate pathways'. It includes a version history section showing 'on: 2 (latest)' and a 'Change to: 2 (latest)' button. The uploader's name 'Jits' and a 'Sign Out' button are visible. The workflow description states: 'Microarray CEL file to candidate pathways' and 'Created on: Wednesday 03 October 2007 @ 18:35:55 (GMT Davlight Time)'. A 'Download Scalable Diagram (SVG)' button is present at the bottom.

Description:
This workflow takes in a <http://www.bioinf.manch.ac.uk/> and returns by this workflow are a list of the top diff on the number specified as input – geneNumber pathways which may be influencing the observed identifying the candidate pathways, more details can be obtained.

CEL
Download Scalable Diagram (SVG)

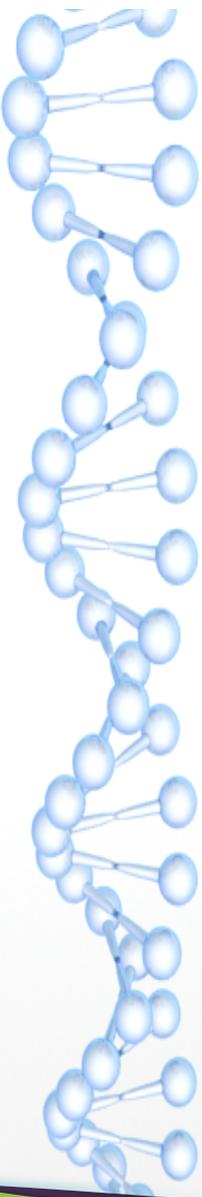
This screenshot shows the Taverna Workbench 2.2.0 interface with annotations. A black arrow points from the 'Workflow explorer' on the left to the 'Workflow diagram' on the right. The 'Workflow diagram' shows a similar flow to the previous screenshot but with different service connections. The 'Workflow explorer' on the right shows a tree view of the workflow components, including 'Workflow9', 'Workflow input ports', 'Workflow output ports', 'abrahamDescriptors', 'Services', and 'Data links'. The 'Data links' section shows connections like 'getMolarVolumeAndRefractivity.volume -> getAbrahamDescriptoi' and 'getMolarVolumeAndRefractivity.refractivity -> getAbrahamDescrij'.

Intermedios: Reutilizan y modifican los SFTC existentes



Tipos de usuarios

No experimentados: ejecutan flujos de trabajo en interfaces simplificadas



Proyecto Inicio Flujos de Trabajo Ejecuciones Acerca de Contacto Iniciar sesión / Registrarse

Foto tomada por Abraham Nieva

Bienvenido al Laboratorio Virtual de Biotecnología de la UPP
 Para soporte técnico y preguntas acerca del Proyecto LVBUPP, por favor visite [Página de contacto](#).

Tipos de análisis disponibles

- Taxonomic Refinement
- Ecological Niche Modelling
- Metagenomics
- Phylogenetics
- Population Modelling
- Ecosystem Modelling

CONACYT LVB UPP es patrocinado por la Universidad Politécnica de Puebla y por CONACYT como parte del programa de apoyos para el fortalecimiento del posgrado 2015. Este portal usa/adapta software desarrollado por los proyectos europeos: **BioVeL**, **SEEK** y **Tavema**

Portal version: 1.6.1-4b07e5d



2.1 Beneficios de los Flujos de Trabajo

- La transferencia de datos de una herramienta a otra es automático
- Las diferencias en formatos de entrada/salida se solucionan con servicios locales de procesamiento (*shims*), lo que facilita la combinación de herramientas



2.1 Beneficios de los Flujos de Trabajo

- Los resultados de un flujo de trabajo pueden servir como entrada otros sistemas o módulos de visualización de datos
- Los flujos de trabajo permiten el análisis de grandes cantidades de datos



2.2 Barreras al usar los Flujos de Trabajo

- Los flujos de trabajo son creados por expertos en TI, generalmente bio-informáticos
- Requieren la instalación y personalización del software
- Se diseñan generalmente para un tipo de usuarios, expertos
- La curva de aprendizaje para usuarios no experimentados es muy alta



3. Laboratorio virtual de biotecnología

Propuesta de trabajo:

Contar con un **Laboratorio Virtual de Biotecnología (LVB)**, un conjunto de herramientas y recursos en línea que se integran para facilitar y enriquecer los procesos de investigación que requieren análisis de grandes cantidades de datos



3. Laboratorio virtual de biotecnología

Propósitos:

- Reducir la barrera para nuevos usuarios (no instalar y no configurar software)
- Adaptar y difundir el LVB para facilitar el análisis de datos en estudios de biotecnología



3. Laboratorio virtual de biotecnología

Beneficios principales del LVB:

- Ejecución en línea de flujos de trabajo
- Análisis transparente, repetible y verificable
- Revisión del proceso experimental (información de servicios, resultados intermedios o detalles de ejecución)



3. Laboratorio virtual de biotecnología

Beneficios principales del LVB:

- Mecanismo de anotación avanzada utilizando ontologías para apoyar la formalización del significado de los datos
- Publicación de flujos de trabajo como objetos de investigación



4. Estudios de usabilidad para el LVB

Trabajo actual:

- Adaptar un laboratorio virtual (BioVeL) para su uso por biotecnólogos de la UPPue
- Verificar la usabilidad del LVB



4. Estudios de usabilidad para el LVB

Trabajo actual:

- Investigar el uso de ontologías en anotaciones semánticas
- Existe un modelo semántico establecido para creación de objetos de investigación basados en flujos de trabajo. Este modelo permite anotar las partes del flujo de trabajo y agregar información adicional
- Se determinará si los objetos de investigación son útiles para: difusión, entrenamiento y reutilización



4. Estudios de usabilidad para el LVB

Trabajo actual:

-Adaptar y crear nuevos flujos de trabajo para facilitar tareas de análisis de datos

- Dos flujos de trabajo de análisis filogenético
- Un flujo de trabajo de modelos de población
- Un flujo de trabajo para modelar bio-reactores



4. Estudios de usabilidad para el LVB

Herramientas de software:

- R para generar modelos matemáticos
- Taverna (SFTC) para automatizar la ejecución de scripts R
- El portal LVB para publicar y probar las versiones del flujo de trabajo



4. Estudios de usabilidad para el LVB

El LVB ofrece oportunidades de desarrollo como:

Herramienta de investigación

Medio de difusión de resultados

Complemento a publicaciones académicas

Material para enseñanza

Modificar el portal para mejorar la usabilidad (en curso)



Publicación de objetos de investigación:

Los objetos de investigación creados se publicarían en el portal del LVB para:

- Validar el procedimiento experimental
- Reutilizar datos, resultados y métodos de análisis
- Incluir ligas a datos y recursos que enriquecen la investigación

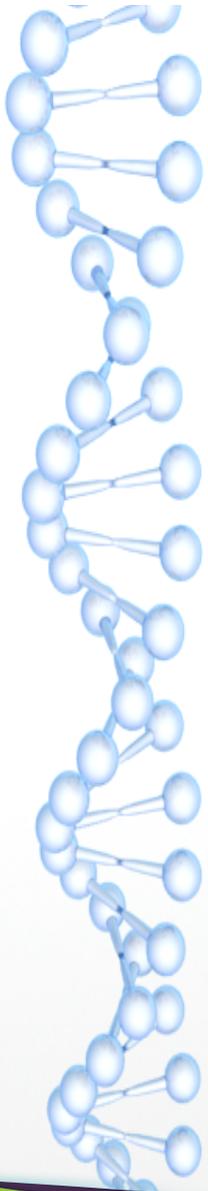
Se espera que así se acelere y simplifiquen los procesos de análisis



4. Estudios de usabilidad para el LVB



LVB reutiliza software diseñado y probado en los proyectos europeos: BioVeL, SEEK y Taverna

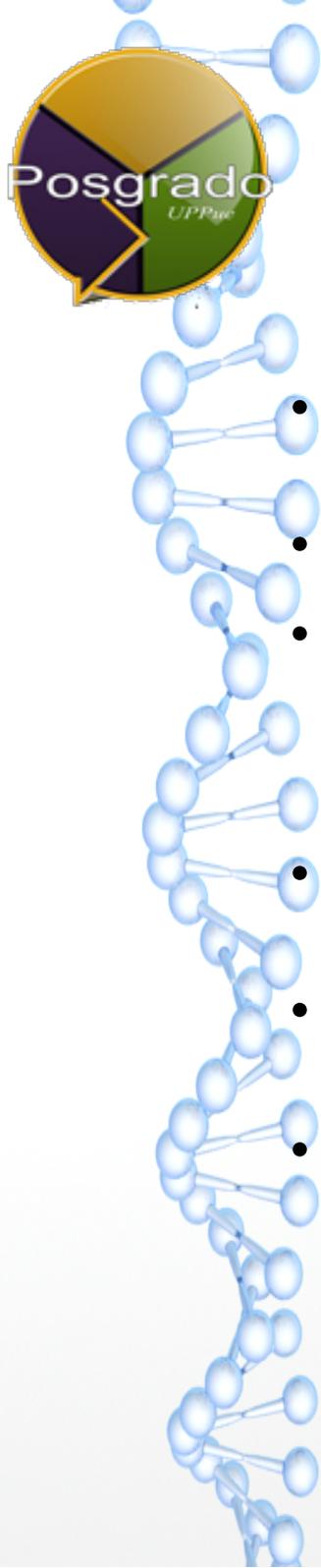




4. Estudios de usabilidad para el LVB

Tareas:

- Evaluar la usabilidad de la plataforma con usuarios no experimentados
- Determinar requerimientos
- Identificar oportunidades de desarrollo como:
 - Herramienta de investigación
 - Medio de difusión de resultados
 - Complemento a publicaciones académicas
 - Material para enseñanza
 - Modificar el portal para mejorar la usabilidad (en curso)



Pruebas de usabilidad

- Laboratorio de Experiencia de Usuario UPPuebla
- Participantes: 10
- Divididos en dos grupos: expertos del dominio (DE) y expertos en tecnologías de la información (TI)
- Tiempo de prueba: ~1/2 hora por participante
- Total de tiempo de pruebas: 5+ horas (agregado)
- Método: Evaluación de sitios web usando el pensamiento en voz alta (Krug, Nielsen)

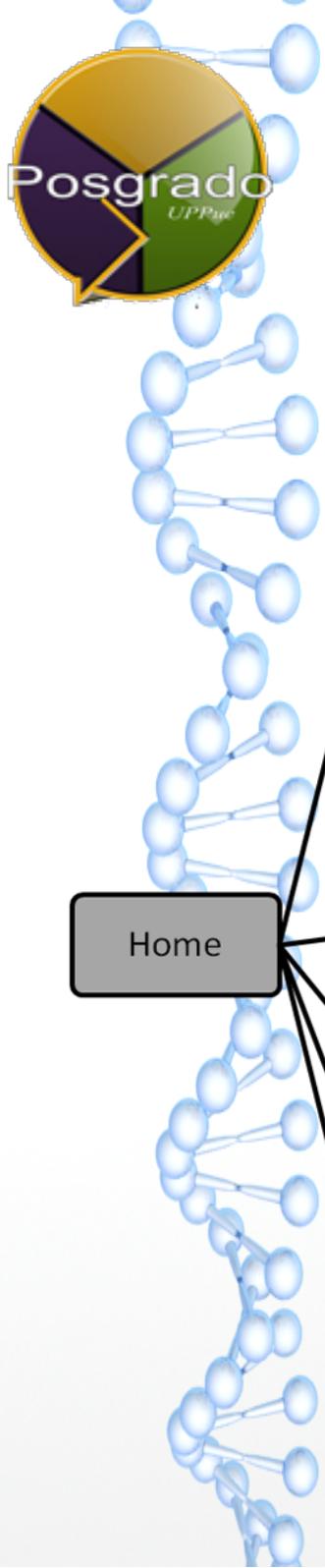
Características	Grupo DE	Grupo TI
Educación	Doctorado: 2 Estudiantes: 3	Estudiantes: 3
Experiencia en análisis de datos	Competente	Limitada
Habilidades en uso de computadora	Usuario regular	Expertos en TI

m@UX

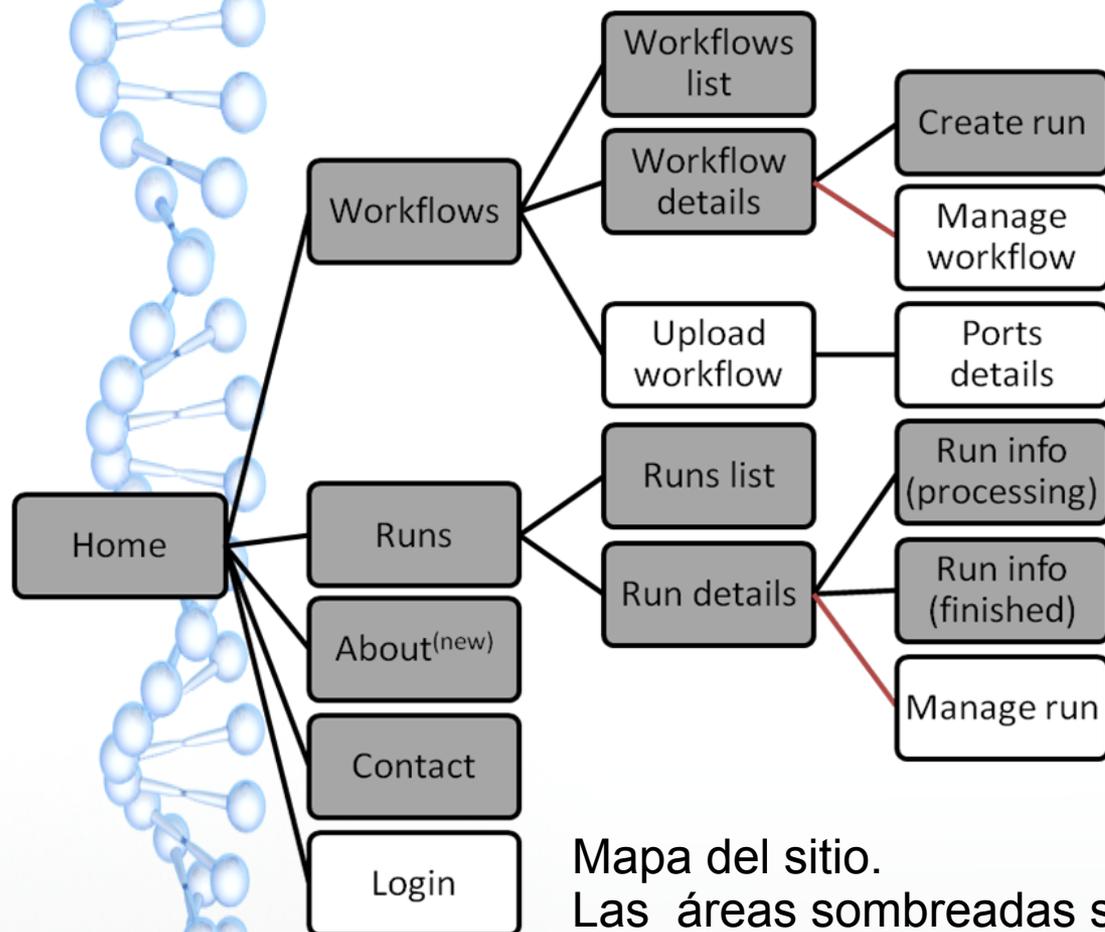


Laboratorio de
experiencia de usuario





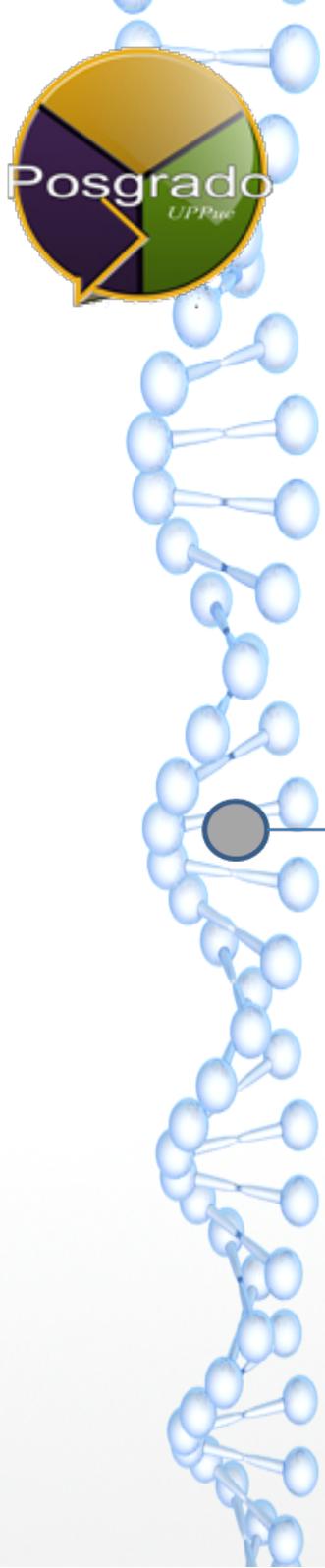
Objetivos de Prueba



- Evaluar

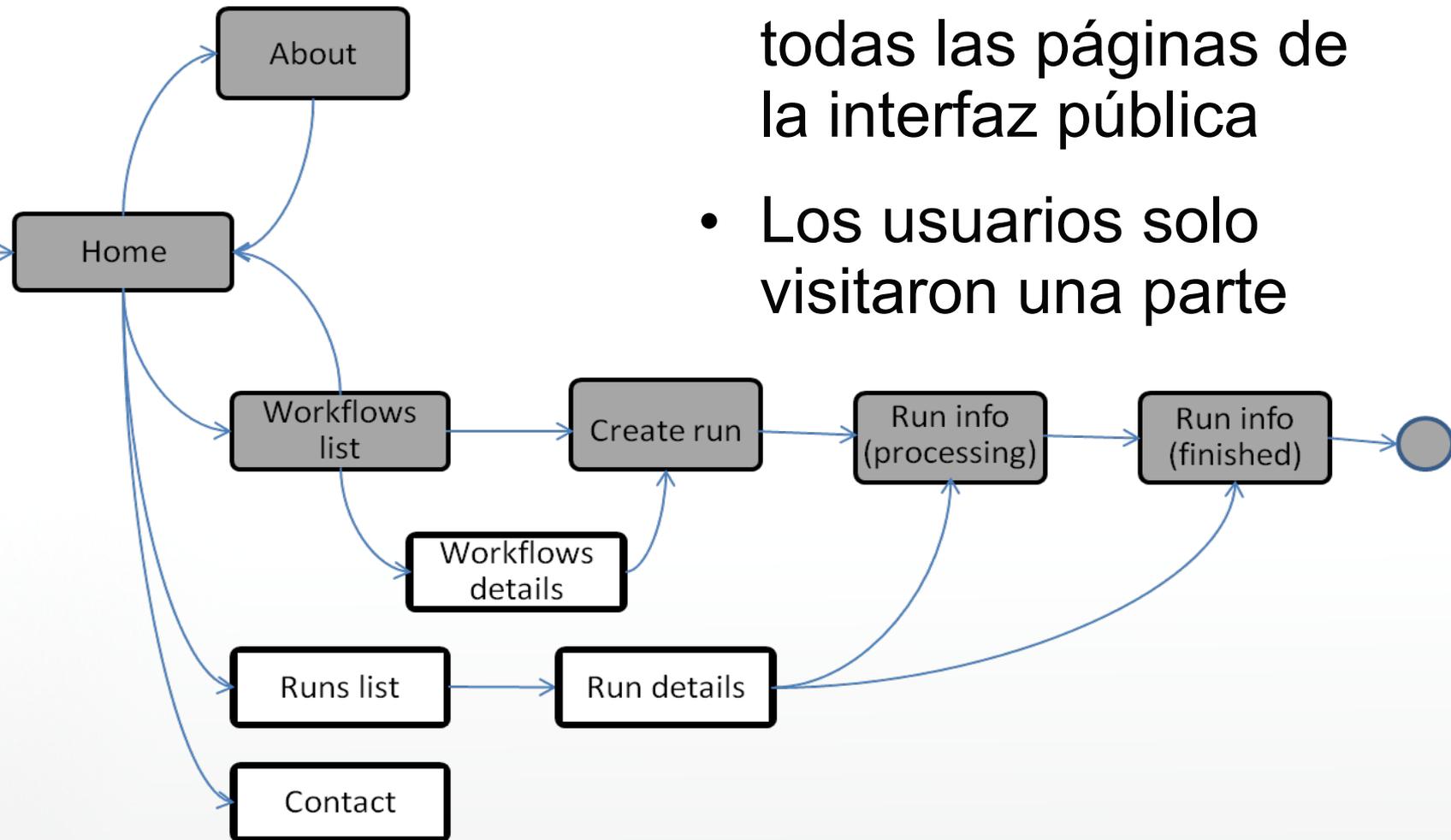
- Interfaz pública del portal
- Navegación
- Información proporcionada
- Complejidad para usuarios nuevos

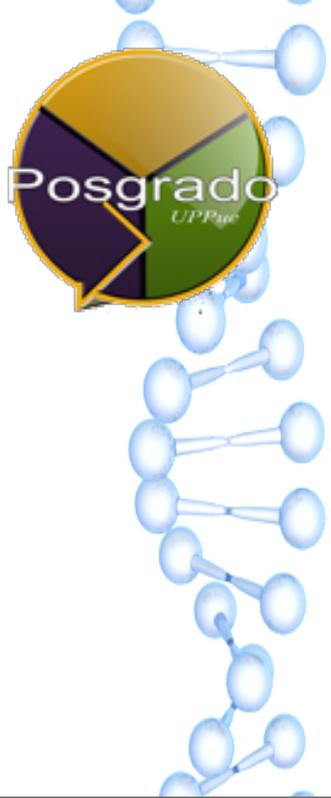




Conducción de la Prueba

- Se esperaba que los usuarios visitaran todas las páginas de la interfaz pública
- Los usuarios solo visitaron una parte

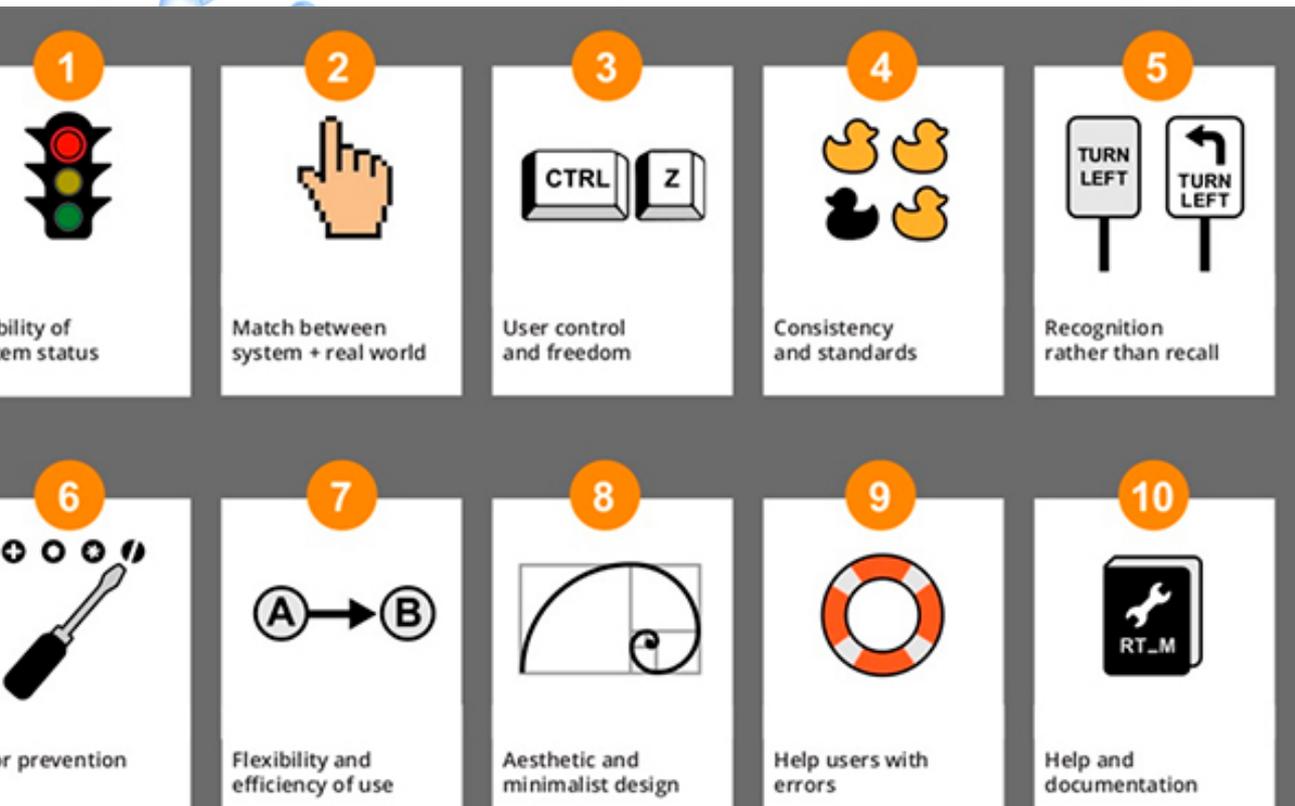




Evaluación de Resultados

- Los problemas fueron clasificados conforme a las heurísticas de usabilidad de Nielsen
- Problemas Detectados:

- Consistencia
- Sobrecarga memoria del usuario
- Ayuda/Documentación insuficiente
- Retroalimentación inadecuada





Mapeo de problemas a requerimientos



Requerimientos	Cosnsistencia	Disminuir Carga Mental al Usuario	Ayuda/ Documentación	Retroalimentación
Definir un conjunto de controles estandar	X	X		X
Establecer reglas de diseño para el uso de controles	X	X		X
Traducción completa	X	X		
Proporcionar acceso a tutoriales			X	
Revisar y Mejorar la interfaz para listar flujos de trabajo <ul style="list-style-type: none"> • Arreglar ligas a detalles del flujo de trabajo. • Arreglar el panel de filtros. 		X	X	
Revisar y Mejorar la interfaz para crear ejecuciones <ul style="list-style-type: none"> • Arreglar los controles de Ejecutar/Cancelar. • Mejorar controles de captura de datos • Mejorar información acerca de tipos de datos 		X	X	
Revisar y Mejora la interfaz para monitorear ejecuciones <ul style="list-style-type: none"> • Resaltar transiciones de estados. • Predecir tiempo de ejecución 		X	X	X



¡Muchas gracias por su atención!

Contacto:

Dra. María Auxilio Medina Nieto

Coordinadora de la Maestría en Gestión e Innovación Tecnológica

mauxmedina@gmail.com

<http://informatica.uppuebla.edu.mx/~mmedina>