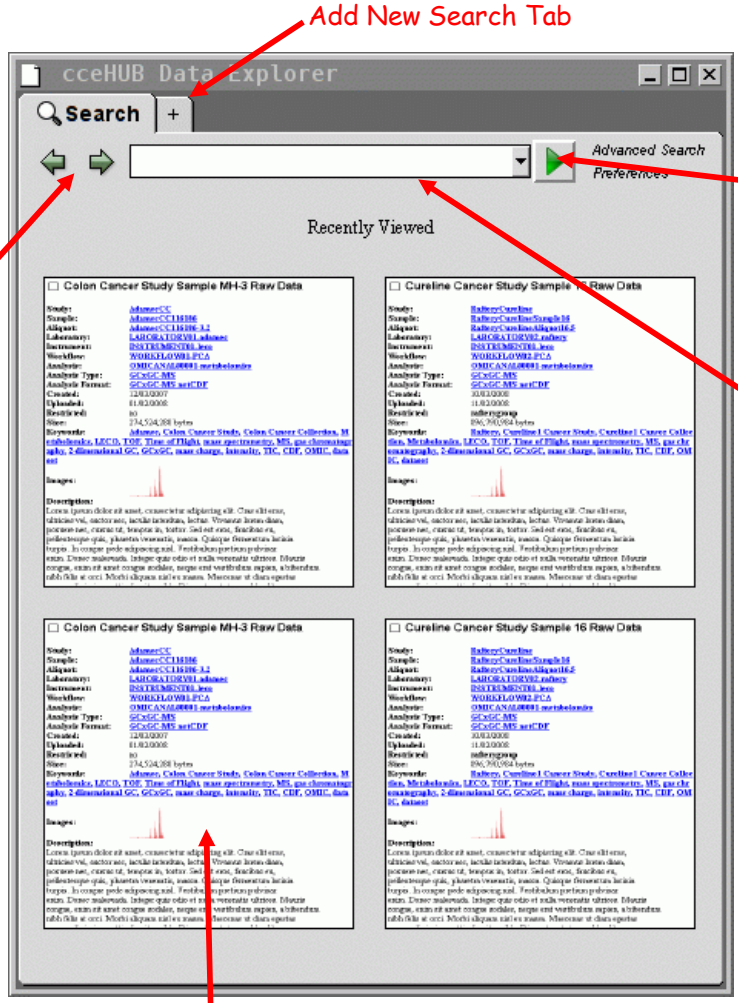
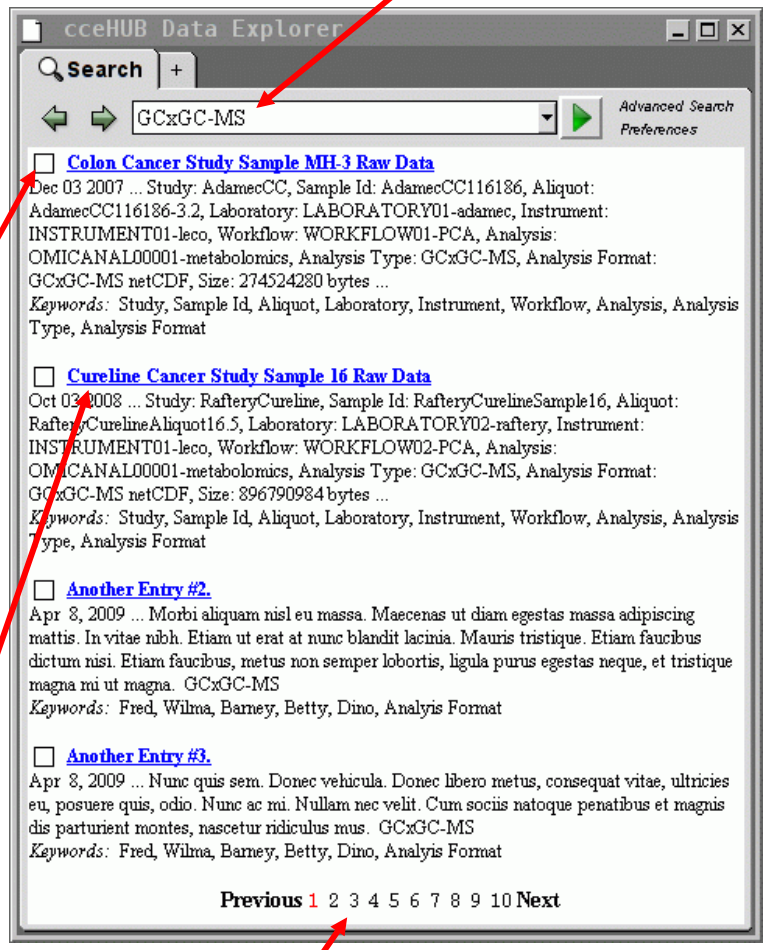
cceHUB Data Explorer

- Problem: Lots of data
  - » Samples, new tool results, etc.
  - » Very large datasets, MBs or GBs.
  - » How to easily select data/results of interest.
- Browser-like data explorer
  - » Search metadata via keywords (tags) and free text.
  - » Leverages metadata and keyword processors.
- Two different applications
  - » Data selector from inside tool (analysis, visualization, etc).
  - » Generalized data browser.



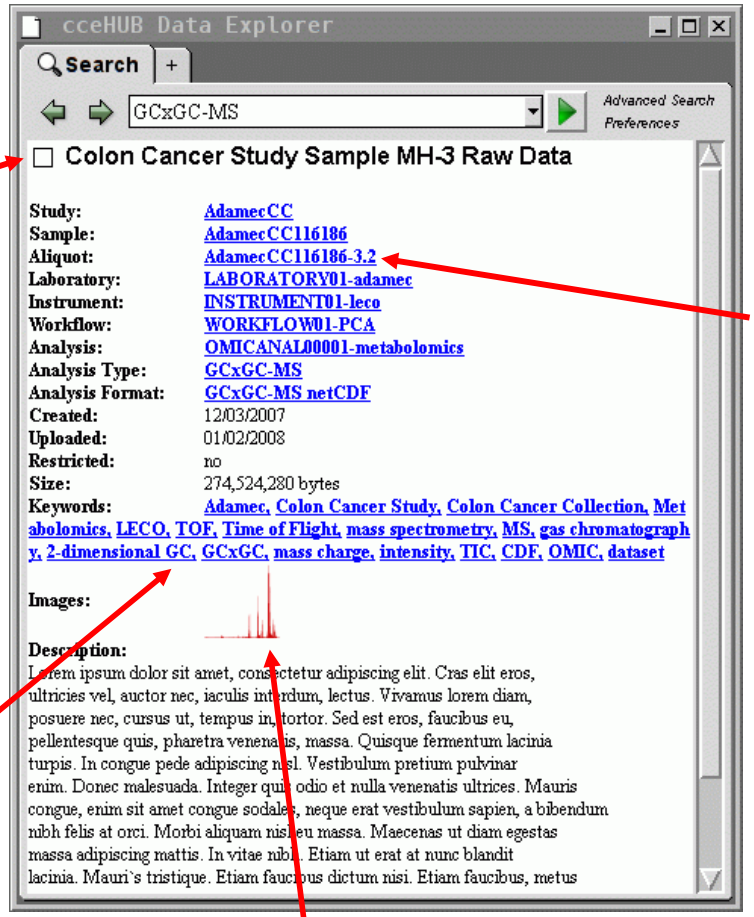
cceHUB Data Explorer

- Google-like search mechanism
  - » Can be context free: “GCxGC-MS” vs. “AnalysisType:GCxGC-MS”
  - » Check boxes to select datasets.
- Search results are ranked
  - » Keyword, full/partial match, etc.
  - » Results also scored by user/group popularity.
- Efficiently queries database
  - » Returns IDs of matches, sorted by score.
  - » Full data content requested on demand.



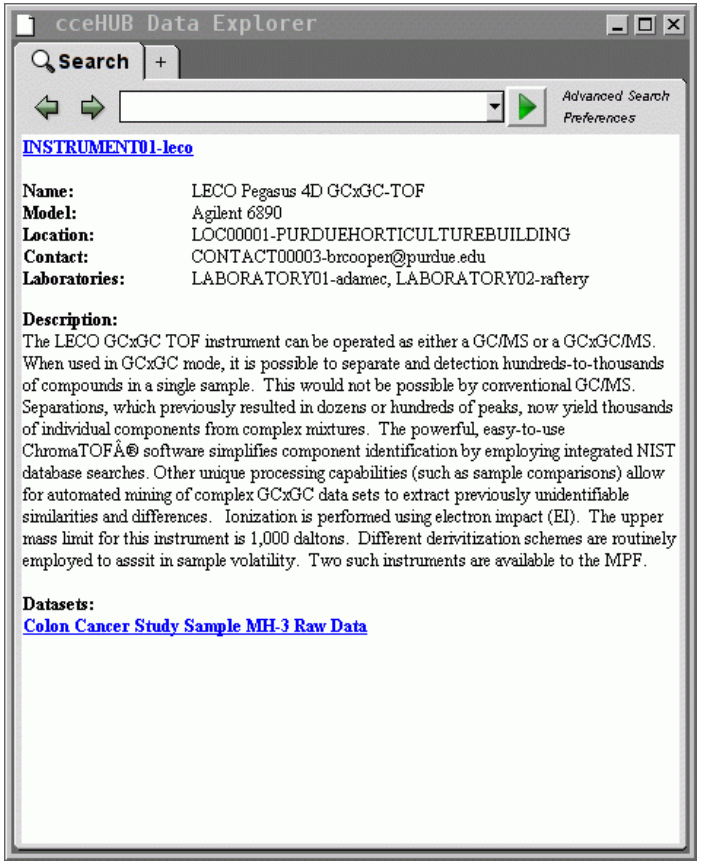
cceHUB Data Explorer

- Entry page contains
  - » Description of dataset.
  - » Check box to select entry.
- But also...
  - » Links to category pages. example: “Instrument”.
  - » Links to tools.
  - » Images (graphs) from tool results.
- Keyword links perform search.
- Leverage Rappture to provide easy connection with tools.



cceHUB Data Explorer

- Content pages
  - » Display more general information.
  - » Describe how data was collected/processed.
  - » Link to similar datasets.
  - » Simple HTML (generated by backend process).



cceHUB Data Explorer

- Google Model
  - » Ranked results (web2.0).
  - » Unstructured data searches.
- Structured data vs. unstructured data
  - » Structured search: query reflects knowledge of database schema.
    - ✓ Good for programs. Range searches. When you know what you want.
  - » Unstructured search: structure imposed on-the-fly as searched.
    - ✓ When you don't know exactly what you are searching for.
  - » Most important point is how easy to find/access your data.
- Data explorer uses both methods.
  - » Displays a table of search results.
  - » Table entries can be sorted and narrowed.
  - » Good for range searches. "*Show entries where the peak retention time is between...*"

cceHUB Data Explorer

- Future plans.
  - » Implement table range searches.
  - » Be able to invoke tools from within data explorer.
  - » Connect to cceHUB database.
  - » Connect to keyword and metadata processors.
  - » Score pages by popularity.