# Indiana University Data Science Knowledge Base Requirements

### Purpose of the Site

1. Provide a resource for students (from Indiana and world-wide in the future) to acquire DS knowledge rooted in actual usage experiences.
2. Become an asset in the broader Data Science community by showcasing IU work.
3. Provide a forum for IU students to share their work. (could be even outside IU in future, but we focus on IU for now)
4. Connect other related IU knowledge, such as the Youtube channel, and ongoing research areas.

### Structure and Requirements (current best thinking, will evolve over time) ***Data Structure***

1. Tools (Specific tools that can be applied)
   1. Tool name
   2. Tool Logo image
   3. Description
   4. Open Source | Commercial
   5. Home Page URL
   6. Student or Trial Usage offer URL
   7. Description of offer (short text)
   8. Related Concepts (zero or more linkages)
   9. Related Solutions (zero or more linkages)
2. Concepts (data science technique, process, technology, algorithm)
   1. Concept Name
   2. Concept image
   3. Category
   4. Subcategory
   5. Description
   6. When to use it
   7. When not to use it
   8. URL (possibly a few URLs to places where the Concept is defined in detail, such as Wikipedia)
   9. Related Tools (zero or more linkages)
   10. Learning Resources (links to OnRamps)
3. Solutions a.k.a Project Experiences (Write-ups that are tagged as follows)
   1. Uploaded Document
   2. Document Preview images
   3. Solution’s abstract (short summary, 1-3 sentences)
   4. Solution Type (Case study, Research Paper, Student Contribution)
   5. Format (Document, Video, e-learning, )
   6. URL to document or video or…
   7. Difficulty
   8. Authors (could include city/country – may want to put in own structure)
   9. Publication Date of the article
   10. Related Courses
   11. Related Tools Used (zero or more linkages)
   12. Related Concepts (zero or more linkages)
   13. Data Sources (zero or more linkages)
4. Data Sources
   1. Possible addition for later.
5. Authors
   1. Future add, to separate authors as a dimension so users can see information grouped by Author, and to ensure Authors are represented only once in the data for proper data normalization.

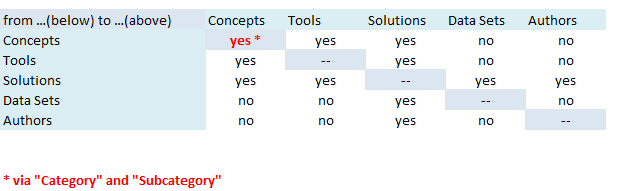
#### General Requirements

1. Search Capabilities
   1. Search should allow for full-text search of uploaded Solutions documents
   2. Allow for filtering via configured search facets
   3. Each document type (Tools, Concepts, Solutions) can have different search facets
2. Search Results
   1. When searching against Solutions documents, provide the following:
      1. Snippet of matching text
      2. link to full document for download
      3. Show other metadata as configured
      4. Show lists of connected Tools and Concepts with links for direct navigation to those entries
      5. Show a preview images
3. Solutions Documents
   1. Store locally in the site uploaded documents in Microsoft Office (Word, PowerPoint, Excel, etc) formats and Adobe pdf.
4. User Roles
   1. Administrator
      1. Ability to create and delete contributor access for solutions.
      2. Ability to upload latest sheet and have website update with latest information
      3. Ability to edit content and layout in main DSKB page.
   2. Content Contributors
      1. Ability to login and create, save, retrieve and publish a solution page.
      2. Ability to edit the latest Google DSKB sheet.
      3. Ability to work with TaDa team to create the "solutions" template
   3. General Users
      1. Can search across all uploaded content
      2. And register in a related forum discussion. (could be via a separate threaded discussion tool)
5. Audit Tracking
   1. Maintain a web log of site page visits and searches performed
   2. Allow interfacing of audit log to Google Analytics for usage analysis.

6. Navigation Capabilities

It should be always possible to navigate between all sections resp. “bubbles” (i.e. Solutions, Concepts, Tools, Datasets and possibly Authors) forwards and backwards, i.e. in any (meaningful) direction

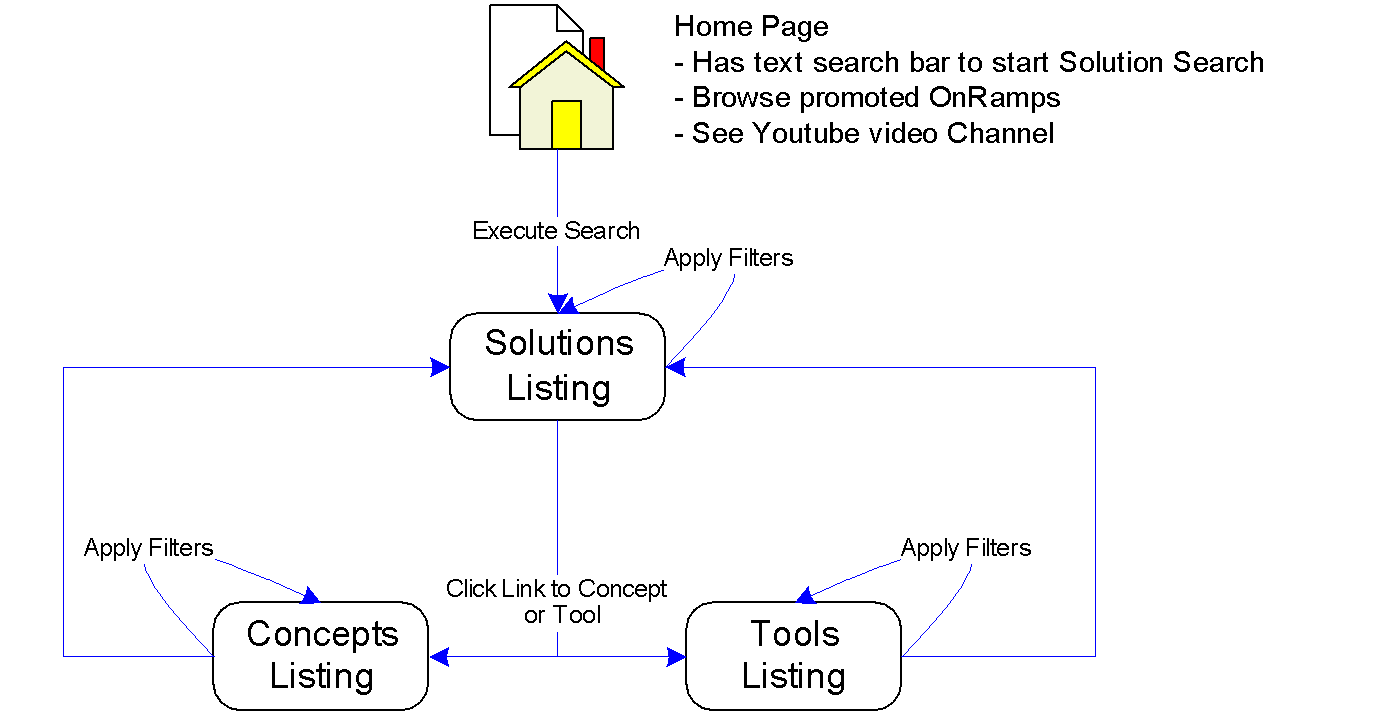
To illustrate this, the following navigation matrix may help:



### Use Cases

1. Home Page
   1. User is presented with engaging home page with Search Bar, site logo and theme, announcements on changes/new content/news. The home page can also house other content such as the IU Youtube channel within a frame.
   2. User can perform text search on “Solutions” and proceed to search results Page.
2. Solutions Search Results Page
   1. Can review solution
   2. Click to visit concept page, or tools page
   3. Apply Filters based on assigned metadata to result set.
3. Pages for Concepts and Tools would be very similar:
   1. When entered from a link on search page, the selected item is a top
   2. Auto-filtering is done to show all other items with matching filter values.
      1. Focus First on Tools and Projects
      2. Can we store documents directly on tada labs site? Would like to be able to use as repository.

### Page Flow



### Mock-Ups

**See ppt from Michael**

### Scalability Requirements

1. Plan for 1000 Solutions documents and 500 Concepts and 250 Tools entries
2. Support 20-30 concurrent users searching site. This is based on 1-2% of Indiana’s data science program. Allow for growth beyond this when site is becomes known outside IU.