International Development Thesaurus – Moving Collaboration and Assessment Upstream

Dr. Denise A. D. Bedford
Goodyear Professor of Knowledge Management
Kent State University
Kent OH

Presentation Overview

- The Challenge
- The Idea
- The Plan
- The Construction (Assessing to understand)
- The Evaluation (For improvement and comparison)

The Challenge

Many voices, many views

- Economic development community represents a wide range of knowledge domains.

- Currently represented by research experts, project teams, consultants, policy experts, etc. from at least 30 high level knowledge domains

- Represents multiple perspectives --
  - for profit, not-for-profit, non-profit
  - International, regional, national, local
  - Emerging as well as long standing issues

Sharing development information and knowledge

- For years the community has recognized the need to improve its knowledge access, transfer, exchange, reuse,…. but the challenge of *uncommon* vocabularies prevents this from happening

- While the subject matter experts may speak a common language when they collaborate or exchange ideas, that commonality is not always evident in organizational vocabularies

- Variations derive not so much from different perspectives (we don't see this in the knowledge communities) but in the structures and warrants of the vocabulary systems

The current challenge

- We have a wealth of vocabularies – at all levels, for all domains, reflecting many different business views

- Vocabularies reflect process in place – harmonization is "end of process" or an "add on to the process itself"

- Many attempts to harmonize, develop mappings and crosswalks – none have proved to be practical or

- Any attempt to harmonize is after the fact – at the end of the process, where structures and concepts are operationalized – impossible to change

- End of process harmonization requires a lot of reverse engineering and is never going to provide sufficient value to a single institution to make the investment - not sustainable

The current challenge

- Generally, there is a small group of individuals devoted to vocabulary development – in some cases it may be as little as one person or a part of one person

- Development of vocabularies requires a high initial investment - investments in people, in supporting technology, in products and integration with existing infrastructure

- No opportunity to leverage economies of scale

- When budgets are cut or people retire, there is a significant risk of knowledge transfer – many vocabularies simply cease to exist

- These efforts will not scale without a larger supporting community – and we need these efforts to scale because the future is semantic

The Idea

Start at the beginning and evolve

- Create an environment in which the community of organizations and individuals who are working in any area or aspect of economic development can collaborate to craft a common vocabulary

- Collaborative environment that allows participants to create and sustain a concept bank here collaborators can deposit concepts and withdraw them as needed

- Allows participants to frame the issues and negotiate meaning at the beginning of the process rather than at the end

- Where governance is loosely defined at the community level, enabling participants to deposit/withdraw to support their own organizations
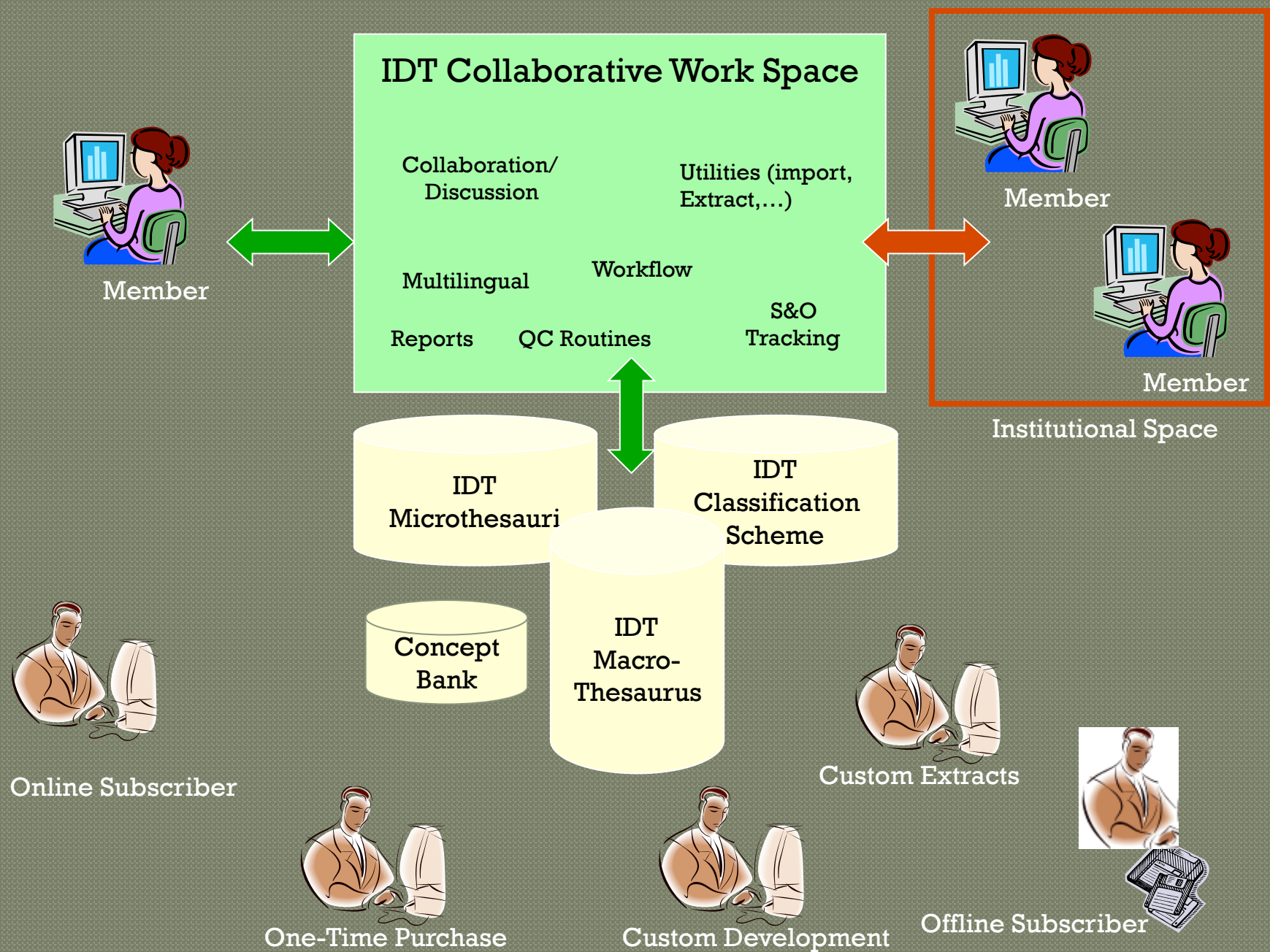
What is it?

- The International Development Thesaurus will be a key tool for classification and indexing of development related content – including documents, reports, projects, publications, websites, and expertise.

- It's design includes both a macrothesaurus, and many microthesauri

- At the base of the design is a concept bank that is created by the contributions of member collaborators

- We believe that a collaboratively developed and negotiated International Development Thesaurus will help to improve cross-organizational access to and discovery of economic development information

- By a common vocabulary we mean generally one that covers all of the issues of import to economic development and represents all of the different views and interests regardless of whether they are used by all or most of the organizations…

- Evolves from within the community, promotes common areas of understanding, and maintains both a current and a historical perspective – surfaces and exposes variations as they occur and provides opportunities to reconcile "in stream"

- Has a common overarching structure – consistency at the top level and supports flexibility and variation at the lower levels

What it is

- Identifies and evolves a core set of concepts that are relevant to many or all organizations involved in the field

- Is consistent in its underlying structures and levels of granularity

- Allows and encourages variations in perspectives and relationships to provide a more holistic view of the economic development environment

- Enables connections across subdomains to create or expose relationships and linkages – goes beyond a single domain

- We do NOT mean "one" single harmonized vocabulary where everyone agrees

  - Harmonization and variation will evolve in the middle – anchored in a common umbrella structure and elaborated in concepts

  - Through collaborative processes – day by day – a common vocabulary and a common understanding will emerge

  - Through collaborative processes a common understanding of concepts will also develop

Collaborative Workspace

- Providing a collaborative workspace will enable people to see others work

- Will provide the opportunity to discourse about concepts, granularity, alignment, etc.

- Social space allows participants to have conversations about the work – where others can actively participate or lurk/learn

- Enables smaller organizations, or organizations with no resources to support this type of a resource to participate – to let their voices and viewpoints be heard

Concept Bank

- Overarching structure and development classification scheme

- Enables organizations to deposit what they have and draw out new concepts that exist in the Bank

- Supported by a flexible data model which allows multiple relationships to exist for a concept

- Also envisioning the data model includes metadata for the relationships – i.e., who has created it, when, what do they mean by it

- At least standard ANSI/NISO relationships but support a more granular breakdowns where definitions and specifications are provided

International Development Umbrella Topics

Agriculture	Communities & Human Development	Conflict & Development	Culture & Development	Education
Energy	Environment	Finance & Financial Sector Development	Gender	Governance
Health, Nutrition & Population	Industry	Information & Communications Technologies	Law & Development	Macroeconomics & Economic Growth
Poverty Reduction	Private Sector Development	Public Sector Development	Rural Development	Science & Technology Development
Social Development	Social Protections & Labor	Transport	Urban Development	Water

The Plan

The Plan

- The intent is to begin with a foundation which covers as many of the knowledge domains as possible – expect to license the World Bank's Thesaurus to jump start the foundation

- As other institutions contribute their KOS, they will be synthesized – knowledge domain by knowledge domain – we will grow the vocabulary

- We anticipate two curation processes –
    - one process which more closely resembles a synthesis and integration (largely driven by owners and IDT support team)
    - One which resembles collaboration among all of the participants

- Governance principles will be established but will be very loose – the intent is for the common vocabulary to emerge not be directed

The Plan

- The International Development Thesaurus is a living, dynamic product that will represent the current and evolving view of the knowledge of economic development.

- The intent is to actively recruit partners, to leverage graduate research students, and to work with semantic technologies to keep the IDT content up to date and growing year after year.

- The partnership model will ensure that participating organizations have cost effective and persistent access to an economic development thesaurus over the long term.

The Construction

Assessment for Understanding

Synthesis and Assessment

- The initial phase of building the IDT, or the first step for any new member, will be to integrate their knowledge organization systems into the IDT architecture

- We realize that we will be dealing with many different structures, warrants, and principles -

- Assessments will be undertaken for each new incoming resource - assessments for the purpose of integration at two levels – top level alignment and leaf node inclusion

- Is there an assessment framework that we can use to get a big picture of the KOS before we try to synthesize it?

A Framework for Assessment

- What are the dimensions along which we need to understand a KOS before we can bring it into the IDT environment?

- We need a neutral, well defined, quantifiable, multidimensional framework against which any 'thing' that any one is calling a knowledge organization system can be understood

- Anyone who has anything they're calling a knowledge organization system that is used for economic development should be able to synthesize it into IDT

Dimensions of a Knowledge Organization System

- Four essential aspects or dimensions of a Knowledge Organization System

  - Concepts – the nature of the content or values that are delivered or accessed through the KOS such as type, granularity, etc.

  - Relationships – nature, type, extent, specification of relationships, logic associated with relationships

  - Context – the context for which the KOS was developed and in which it may be used, including knowledge domain, application domain,

  - Governance – control and management of the concepts, relationships and context exercised by the developer or current user

Defining Components and Quantifying Factors

- For each of the four components, an orthogonal, independent set of factors must be defined

- Factors must be independent of any particular pre-existing KOS (neutral)

- Each factor must have a quantifiable method of representation that lends itself to 'scoring', analysis and comparison

- Factors must have agreed upon definitions, be easily interpreted by people and machines, and be inclusive in their coverage of values/conditions

- To illustrate the idea, selected examples are presented in following slides

Selected Examples of Concept Factors

- Concept types
  - Data/numbers
  - Calculation/ratios
  - Words
  - Grammatical fragment
  - Logical statement
  - Rule expression
  - Engineering equations

- Degree of ambiguity
  - Context sensitivity/ insensitivity of definition

- Representational form
  - Usable encoding method
  - Availability of representational specifications (Strings vs. syntax)

- Degree of conceptualization/ specification
  - Theoretical to commital

- *What else…?*

Selected Examples of Relationship Factors

- Simple expressive form of relationships
  - Grammatical
  - Mathematical
  - Logical

- Relationship behavior
  - Membership dependence
  - Representation or instance
  - Equivalence
  - Causal dependence
  - Derivational dependence

- Degree of Relationship Validation/Rigor
  - Fully Subjective
  - Grammatical validation
  - Mathematical validation
  - Logical rigor/validation

- *What else?*

Selected Examples of Context Factors

- Knowledge Context
  - Formal vs. informal knowledge domain

- Application Context
  - System vs. human application/ consumption
  - Managed/standardized application vs. home grown

- Functional context
  - Search
  - Mathematical or statistical analysis
  - Logical inference
  - Classification
  - Dynamic clustering
  - Metadata representation
  - Concept indexing

- *What else…?*

Selected Examples of Governance Factors

- Standards Availability
  - Published formal vs. guidelines vs. ad hoc concepts
  - Published formal vs. guidelines vs. ad hoc relationships

- Prescriptive vs. Descriptive Governance
  - Enforcement of standards

- Design Guidelines
  - Top-down (model) vs. Bottom-up (empirical)

- Extensibility
  - Degree to which others can add to or extend either the concepts or the relationships

- Currency
  - Degree to which the concepts and/or relationships represent our current view or knowledge of the context

- *What else…?*

Evaluation for Improvement and Comparison

Factor Analysis for Evaluation

- This same approach may be used to evaluate or characterize KOS in general – for the purpose of

  - Informing users about a KOS
  - Providing developers with methodology for comparison and improvement
  - Identifying minimum standards for what is/is not a KOS
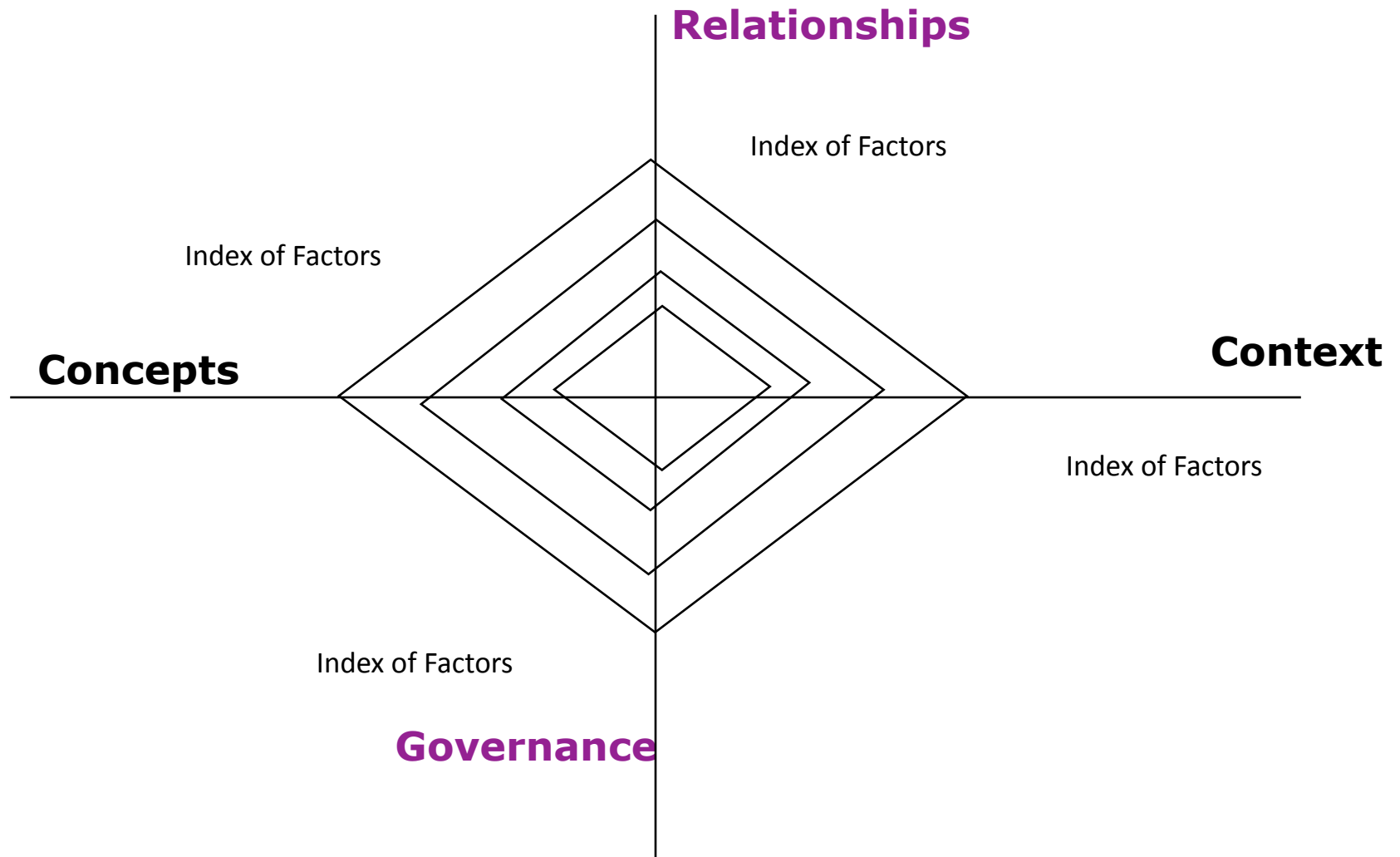 - Thresholds for formal and informal KOS

Factor Analysis for Evaluation

- Factor analysis could be conducted:
    - For any of the four components in the model
    - At the KOS level, and across all four components

- Developers or users could determine what the optimal dimensionality was for their particular use

- Statistical method used to describe variability of factors in which the factors are modeled as linear combinations.

- A single factor in the model would represent a set of 'like' variables which otherwise would be too complex to model
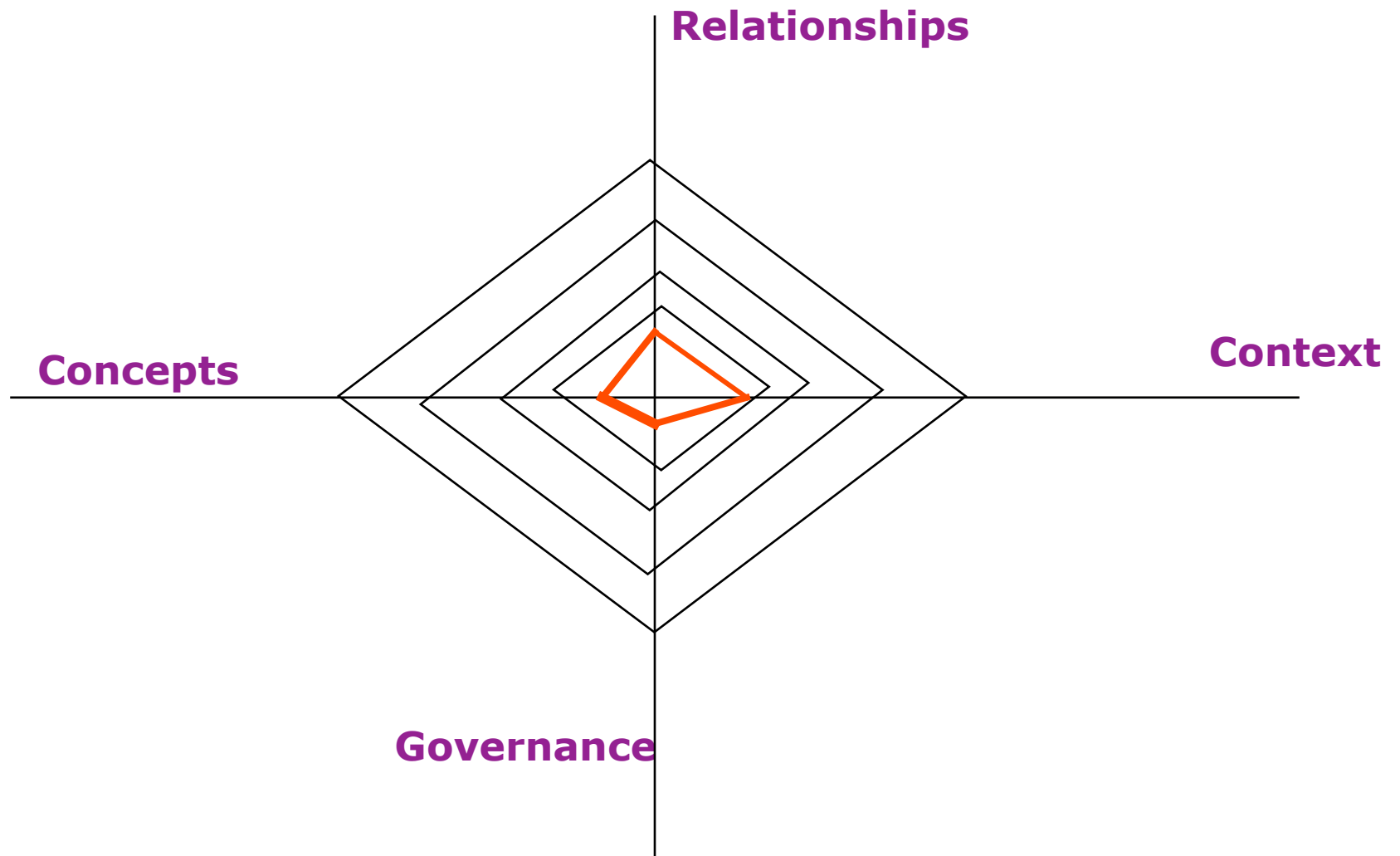
Representation of KOS Assessments

Relationships

Index of Factors

Index of Factors

Concepts

Context

Index of Factors

Index of Factors

Governance

Methodology could be used to generate a factor index for KOS, to rank and compare knowledge organization systems (KOS).
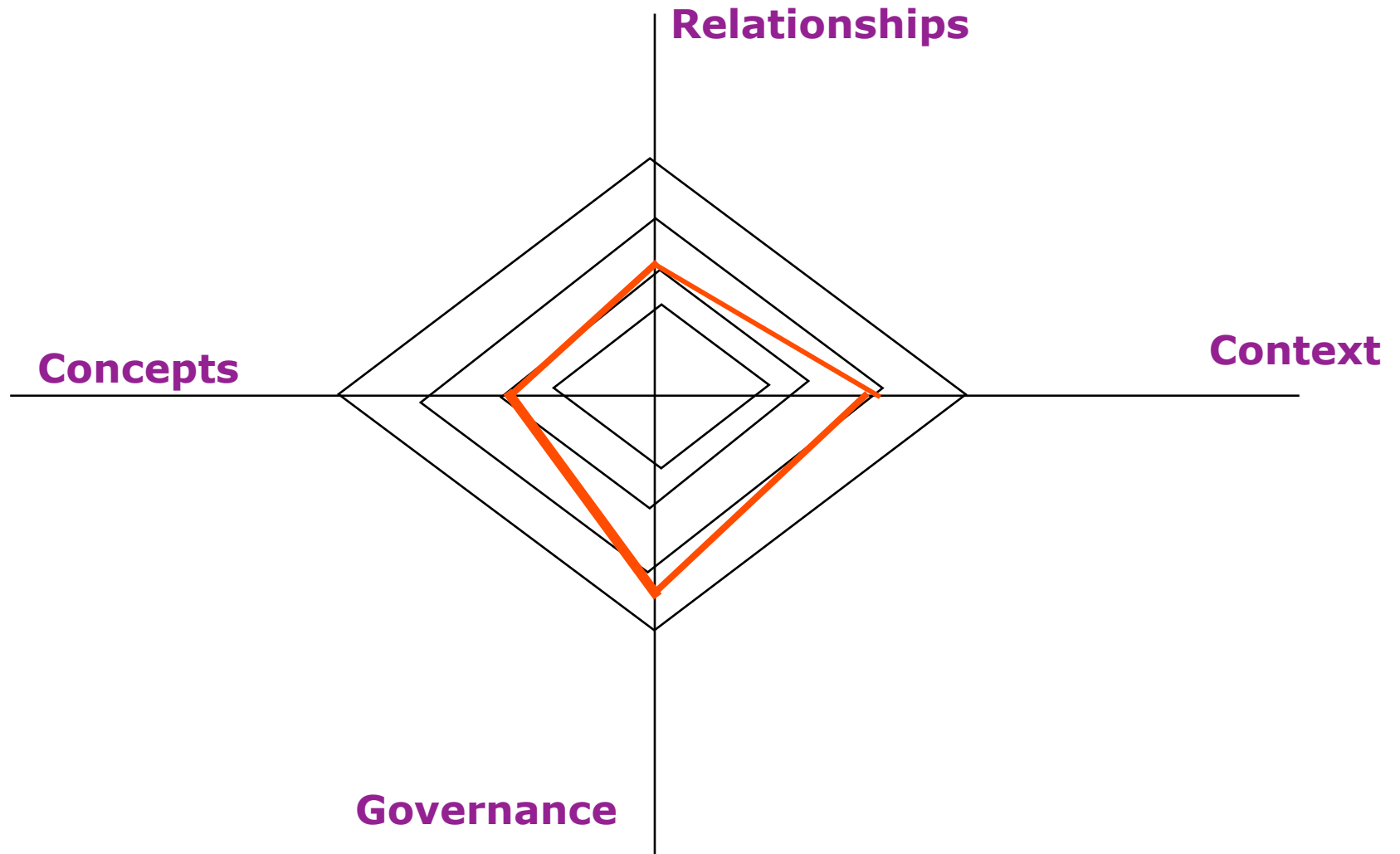
Representation of KOS Assessments
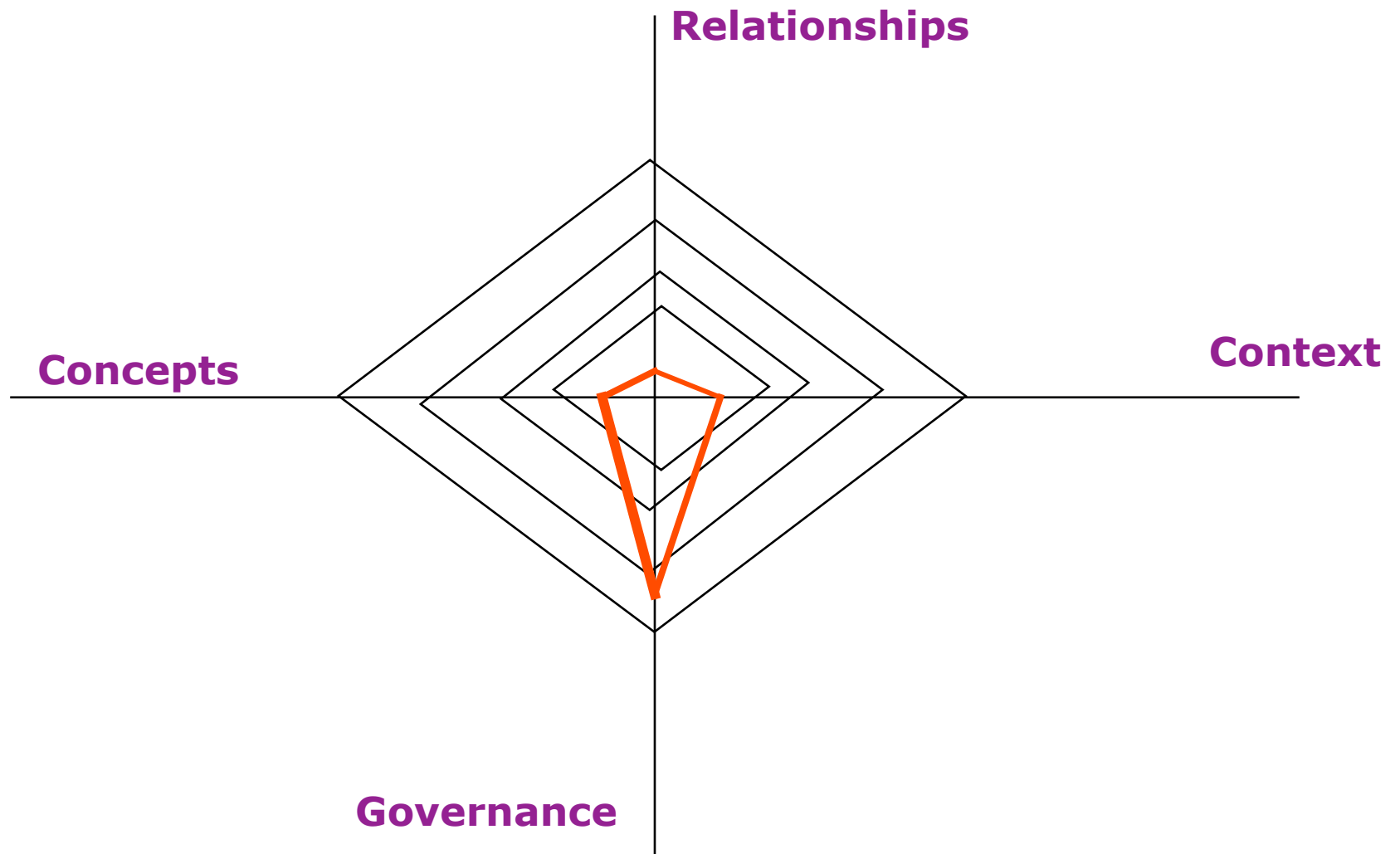
Sample assessment of a *folksonomy*

Relationships

Context

Concepts

Governance

Representation of KOS Assessments

Sample assessment of a *medical disease classification scheme*

Representation of KOS Assessments

Sample assessment of an *institutional records classification scheme*

Observations and Notes

- This methodology can be applied to all types of knowledge organization systems from simple controlled lists, to domain specific classification schemes, or fully elaborated thesauri to ontologies with formal logic

- The framework can help us to quickly identify the "type" of KOS we are looking at, understand its components and decide whether it is suited to a given purpose

- The framework can also raise awareness of key components that should be designed into any knowledge organization system

Dbedfor3@kent.Edu

Questions or Discussion?