[Arofan go to the bottom of page 3]

Mission:

* NOTE: distinguish the missions of DDI and the DDI Alliance. See the [Charter and Bylaws](http://www.ddialliance.org/about/about-the-alliance).
* Possible wording:
  + ~~To develop, maintain, and promote world-leading standards and semantic products describing observational data on human activity and their context through a community-based organization~~.  
      
    Expanded descriptions of the mission:
  + ~~The mission of the DDI is to advance high quality, multi-purpose data documentation through technology-independent metadata specifications, protocols, and semantic products that are community developed, maintained, and promoted and that support the global social science research community in producing, using, preserving, discovering, and reusing observational research data in the social, behavioural, and economic sciences. DDI products are intended to facilitate human and machine-actionable research applications.~~
  + ~~The mission of the DDI Alliance to be the official governance body of DDI, established under the terms of an organization Charter and Bylaws that have been approved and signed by Alliance members, is to provide responsible, transparent, sustained, and continuous community oversight of the DDI product line. The DDI Alliance is a global player in the digital research infrastructure community and openly collaborates with other stakeholders interested in solving common problems in documenting data and the research process.~~   
      
    The mission of the DDI Alliance, as the official governing body of DDI, is to oversee and enable responsible, transparent, sustainable, and continuous community development of the DDI product line and its interoperability with international digital research infrastructure. The DDI Alliance openly collaborates and engages with stakeholders working to solve common problems in the documentation, management, discovery, preservation, and reuse of observational data on human activity and its context.
  + This wording doesn’t include “succession plans” but this could be considered part of “responsible oversight.”
  + The DDI Charter uses the host institution as the legal entity for DDI. I don’t believe that DDI Alliance is incorporated.

Guiding Principles:

* DDI products are:
  + Free and publicly available;
  + Protected through registered trademarks, certification marks, and collective marks;
  + Responsive to the documentation requirements in social, behavioural, and economic research;
  + Adaptive to changes in social, behavioural, and economic research;
  + Directed at providing common solutions to data documentation across domains;
  + Relevant to the communities that use it;
  + Developed using transparent and reliable processes;
  + Accessible at persistent locations;
  + Supported for ease of use;
  + Migratable across versions or release changes;
  + Interoperable with other metadata standards.
* What about the internationalization of the the product line? Will it support multilingual applications?

We don’t have a vision or aspiration statement.

VISION : Aspirational statement (10 years from now, DDI will …)

1. DDI Product Line
   1. DDI-Model has undergone three substantive version changes and supports all previously produced DDI-codebook and DDI-lifecycle content by migration or interpretation.
   2. DDI-Model is a machine-actionable service implemented through one or more registries that applications can query for metadata specification details.
   3. DDI-Model supports bindings to the leading information technologies.
   4. Modifications to DDI-Model are determined by community decisions and implemented directly on the appropriate registries.
   5. DDI semantic products are delivered via registries and identifiable through persistent identifiers that allow their easy use in applications.
   6. DDI protocols support distributed digital research infrastructure move metadata across applications delivering the appropriate DDI content in a predictable format.
   7. Development of DDI-codebook and DDI-lifecycle has ended after DDI-Model utilities had proven successful in processing files containing content in these earlier specifications or in migrating the earlier specifications to DDI-Model.
2. The DDI Alliance and DDI Community Development
   1. A sustainable business plan has been in place for five years and will be renewed for another five years. This plan is based on a mixed contribution formula that includes membership contributions and key stakeholder investments to support web services that they operate and are dependent on the ongoing maintenance of DDI products.
   2. The use of DDI products is less dependent on software vendors building DDI capabilities within their products and much more dependent on Web services making calls to DDI registries to support their applications.
3. DDI Alliance and DDI within Global Digital Research Infrastructure
   1. DDI-Model is the leading standard for the discovery layers of research data repositories.
   2. The DDI Alliance participated in the CCSDS development of its architecture concept.
   3. DDI products play an integral role in the discovery, interoperability, and integration of distributed data sources across data repositories, domains, and countries.
   4. DDI Alliance is in a strategic partnership with a major research information management (RIM) stakeholder in the exchange of shared metadata to provide a more complete metadata record for research data.
   5. DDI products are being used in a major deep learning project to train a neural network how to classify social media sites for producing observational research data.
   6. DDI is integrated into the operations of five NSI’s.
   7. Five national research councils now have data policies that identify the use of DDI as a best practice in the collection and management of data on human activities.

George:

Integration of the standards

Common Data Element registries

Expanding the membership to NSOs and data producers

Research council data policies recognizing DDI in their best practice statements

Journal policies including DDI

[DART](https://www.dartstatement.org/) Data Access and Research Transparency

Engagement as the language to use instead of partnerships

Need to make things easier for the researcher

DDI could be a key player in discovery in digital infrastructure

Achim:

Levels : generic and method (e.g., survey-centric)

How do registries and repositories really work and support discovery

Distinction between data capture and data description

* This distinction between survey-centric data description and other data description

What does data lifecycle mean for web scraped data

More distinctions among purpose of the vision; functionality; examples

Repository: database of digital objects

Registry: catalogue of things; have metadata on metadata; which kind of metadata to index; registry catalogues are recommended to be curated

CDE registry - high-quality metadata, would speed up generation of surveys

Registry -

Distinction between survey metadata and data description that could be applied to anything.

High quality metadata is the purpose of a registry.

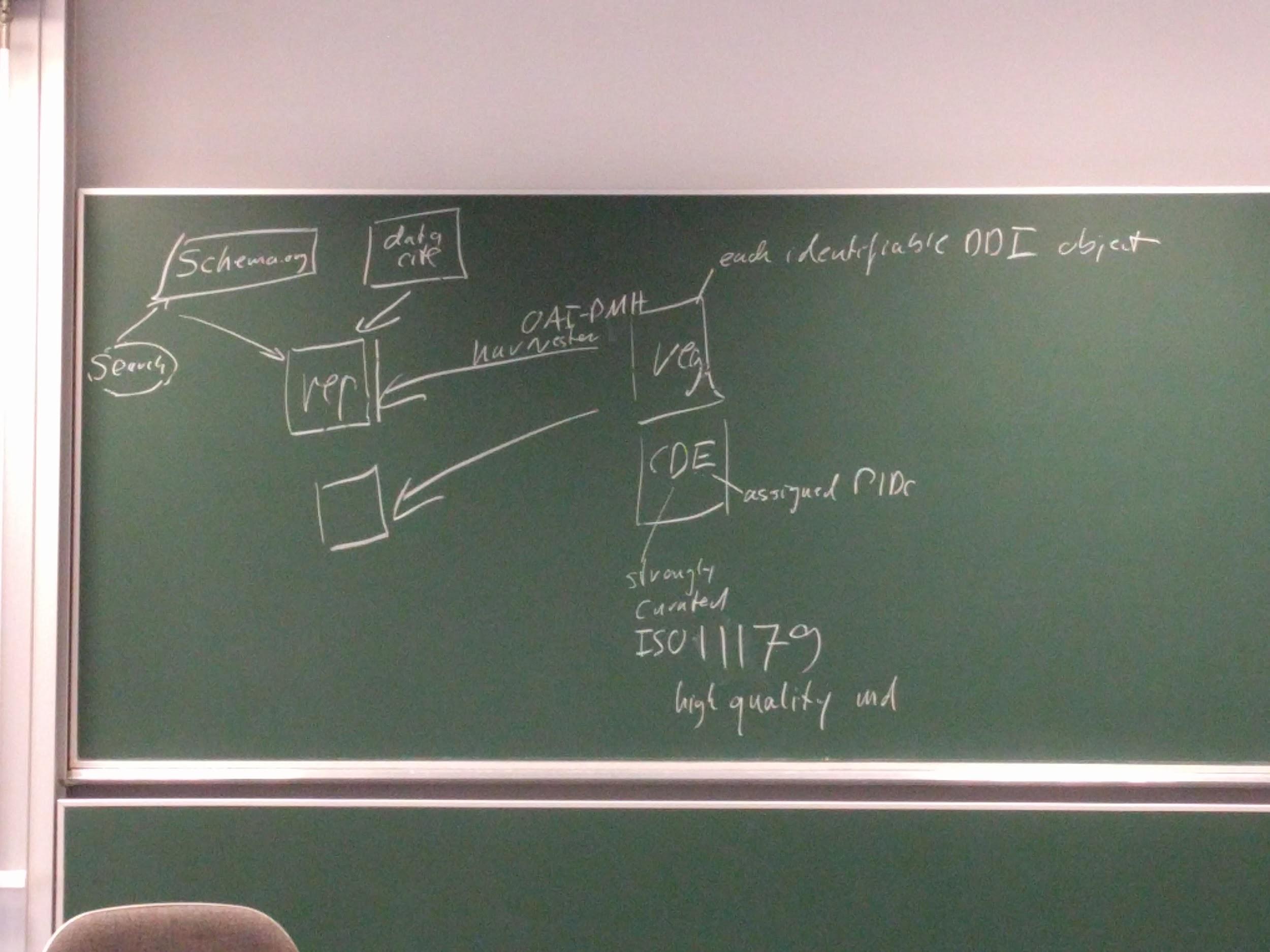
How can the legacy stuff be moved into the CDEs? Which are the most important metadata items that would be candidates for CDEs?

Understanding how we fit into the ecosystem so we can determine how we can play well with others.

What are the parts that the Alliance is providing that is not already part of the ecosystem?

Publish a document showing how to go from DDI 2 (3, or 4) → Schema.org

Chuck: vision includes: web services critical to implementation of DDI model and other specifications, and the interoperability of this within the wider digital research infrastructure. We don’t have to be comprehensive at this time, but we need key strategic services. We have at least two (see chalkboard).



Bill: we also are defining what we are not.

Chuck: In Achim’s diagram, what’s now identified as a repository is actually a large number of repositories.

Achim: search of instances of metadata, and searches for CDEs (representation level)

Steve: When would I use these things? I start with an idea and identifying what do I know about an idea. The next stage is to formalize the idea. Understanding the datasets used.

Chuck: third part of the vision is the process: the process for engaging with the ecosystem.

Steve: we can propose a model and a framework for working together.

Chuck: web services are critical for us moving forward with implementation of DDI model, with a connection to other web services that appear in the wider digital research infrastructure, and the process of how we get these services developed and supported as part of the larger community

Steve: The real question for the Alliance is which parts do we do?

Bill: There’s a marriage between this and the manifesto.

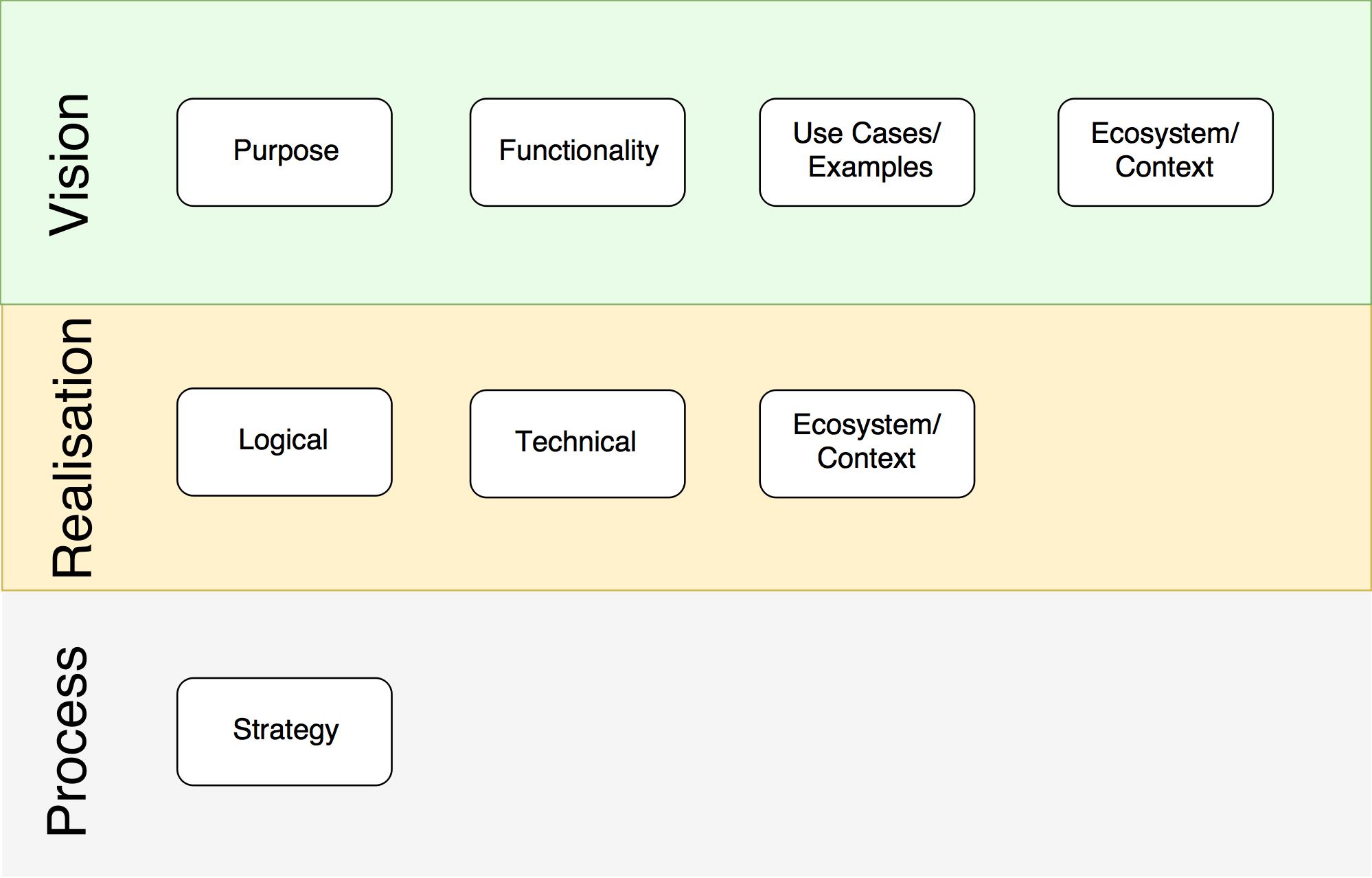
Chuck: What do we want it to be, how we get there, and what are the gaps?

George: Manifesto is the vision, but in order to get there, we need these pieces. We’re not going to do it all ourselves to get these pieces, so we’ll have to engage with others.

Chuck: This leads us to principles. What principles guide that engagement?

George: One is community (not proprietary). Community maintained standards and protocols that are open and not controlled by private entities that claim them as their own.

1. Infrastructure Vision: Purpose and the functionality of the infrastructure
   1. Need diagrams : a lot of diagrams!
2. Realization: logical description, technical
3. Process: strategy



Linking purpose with FAIR and DART

Use cases:

1. Survey design: Development of computer-assisted instruments (CAI). Element registry makes process DDI-based and eliminates some of the bottlenecks. For example, bottleneck between designers and CAI designers. If transfer design in DDI, you eliminate that. So it makes the process better and cheaper.
2. Meta-analysis: you want to find how many studies out there that have these five variables so I can redo this analysis. Being able to link across languages.
3. Interoperability:
   1. You need to be able to take variables from more than one source and create a combined dataset. This gets to the datum level in DDI-Model. Suppose you want to add a contextual variable like unemployment to a dataset at the individual level. Doing this would make it easier to find the variable. Secondly, as repositories expose data in DDI-Model, you could write a script that would create the dataset for you or there could be tools that are specially designed to create that for you. To make it fancier, if using ISS data from multiple countries, adding contextual data for each country, and create a combined dataset.
   2. The exposure of metadata that can enable discovery and data extraction across multiple repositories
4. Access - The exposure of metadata that can enable discovery and data extraction across multiple repositories.
5. Discovery (of concepts, variables, questions, captures, datasets, analyses, publications)
   1. Concept clarity -- registries provide
   2. Pan-repository search
   3. Researcher beginning a project and using the data to find all the publications on that research topic.
   4. What other variables are correlated with a specific measure? Where has this question been asked in a different language? Can I find all instances of this variable, and then access the datasets that include it?
6. Transparency: automate the process of documenting data transformations.
7. Citation: automate the creation of citations for multisource datasets (provenance)
8. Impact:
   1. Demonstrating to research councils the links between funds (grants), researchers, data, publications, and citations.
   2. What is the relationship between this research project and others? Demonstrating where the data and the content get used and reused.
9. Sustainability: One-off, ad-hoc projects that do not get used later. Why don’t the projects get maintained?
10. Harmonization: Harmonizing existing variables from multiple studies into a new dataset. Harmonizing data over time. Harmonizing cross-national studies.
11. Risk Management: Can you tell me the risk potential of this dataset? (E.g., a risk factor for every variable). Can I automate statistical disclosure control? Can I extract my data to the relevant access environment?
12. Citizen Science: How to have citizens participate in the research process?

2017/10/25 AM

SDMX Example

* Solving a common problem: NSO/I repositories sharing
* Federation : technical issue associated with it : example of multiple registries (treated as catalogues pointing to published data residing in repositories) : notifications populated within multiple catalogues : Eurostat rollout is an example for collaboration
* The registry is place where to for data but also to go for metadata
* RDF + XML bindings to information models

Organizational considerations is coordinating collaborations:

1. heterogeneity/homogeneity of services, e.g., repositories
2. Top-down/bottom-up authority for getting things done

With heterogeneous systems, the key are the interfaces to the repositories

Research Data Australia as an example, too

* Repositories at universities harvested and aggregated

Jared:

* What do we do to engage researchers in the use of DDI?

Technical requirements

* Interface and protocol (REST-based : on the level of HTTP: each resource identified by a UI) : two draft proposals for protocols
* Need to look at OAI-PMH to understand if it satisfies needs
  + Jon: suffers from less and less info carried : there is a tipping point in that technology that we need to be aware of : what is the optimal, preferred technology
* First think about requirements and then look at existing technologies
* George: we want use cases that specify the variety of queries we want to be able to ask : describe scenarios and ask what are the requirements and the resulting technologies needed to address these

Client to maintain the registry

Legacy repository could use a protocol : push / pull notification : requirement that the registry learns about new things in the repository

Requirement: identify complex queries

Requirement: human and machine interfaces : allow others to build the human interface on top of the machine interface

Requirement: single registry or a federated registry

What kind of queries would be useful:

* Type of users
  + Start from a concept level : get a number of measures : determine one of interest and ask for data
  + Someone designing a new instrument : look at different measures related to their concept : look at datasets to see if measure really worked

Support analysis models with DDI-Model

* Roles that variables play
* Putting these variables together in an information model
* Create different record types the way put metadata together
* Build out different information models to answer research questions

**Thoughts on repository versus registry:**

A repository stores digital objects. In our case the digital objects are DDI items which are fragments of DDI instances. The metadata repository stores not only the metadata but also adds relationships with related metadata. The metadata repository uses the persistence identifiers which are assigned in the process of the creation of the metadata.

A registry is a catalogue on digital objects. The registry stores metadata on digital objects stored in repositories. It is a little confusing in our case that there is metadata on metadata objects (DDI items). The metadata in the registry serve the purpose to find the right digital objects which are the DDI items. The metadata in the registry reference the DDI items stored in repositories. In this sense this type of registry is loosely related to the concept of a portal. A subset of the information in the repositories is stored in the registry.

This type of registry can hold information on multiple repositories on any DDI items (any identifiable DDI object).

The repositories require a standardized interface to expose the content, for example a REST-based protocol which assigns an URL to any resource i.e. DDI item or group of DDI items.

The registry can get content from the repositories either by harvesting the repositories or by notification from the repositories. The first approach has the advantage that repositories are only required to have the standardized interface. The latter approach requires additional software on the side of metadata repositories side which needs more governance and maintenance.

A metadata registry holds information on metadata for the purpose of definition and exchange. The metadata definitions are stored and maintained in a controlled method. The metadata registry assigns persistent identifiers to the metadata definitions.

The metadata definitions are more on the conceptual side of DDI, i.e. Concept, RepresentedVariable but not InstanceVariable.

This type of registry borrows ideas from ISO 11179 and Common Data Element.

The registry needs input from providers of common data elements who maintain these. The registry maintenance makes sure that the common data elements are imported in a standardized and validated manner.

NIH definition of CDE repository:The Repository is a platform for identifying related data elements in use across diverse areas, for harmonizing data elements, and for linking CDEs to other existing standards and terminologies. This definition doesn’t use the term registry at all.

See: <https://www.nlm.nih.gov/cde/>

NIH page on Guidance to Encourage the Use of CDEs: https://www.nlm.nih.gov/cde/policyinformation.html

Both types of registries complement each other. The first type (portal type) provides access to any DDI-based metadata in public repositories. It would serve current needs. The second type (common metadata element type) serves forward looking needs and provides high quality of metadata.

The kind of metadata of the portal registry is maintained by institutions for different purposes, in a variety of quality (from not maintained to strongly curated). It is a mix of metadata on the conceptual and instance level. It comprehends metadata (on the conceptual level) which could be candidates for curated data elements in the second type of registry (common data element registry).

An “editor team” of common data elements could look for these candidates in the portal type of registry. In a second step it would be required to communicate with the institutions where this kind of metadata is coming from. The possible goal would be to move this kind of metadata from the stage “used” to “curated”. Then they could be included in the common metadata element registry.