DDI Longterm Infrastructure Manifesto: Outline as of 25 October, 2016

1. Overview (Jared, Katja)
	1. Motivation
	2. Purpose
	3. Benefits

Benefits to achieve in the infrastructure system: (& other goals)

* Automate capture of metadata (reduced costs and time)
* Capture better/more complete metadata
* Enable new data discovery and analysis tools
* New data harmonization, comparison and combination tools
* Systems that can be used/across organizations; Transparency across organizations across or w/in stages of the life cycle
* Encourage interoperability and comparability across studies, domains, and countries
* Infrastructure for small scale w/o benefits to overhead
* Lower cost of using/entry into DDI infrastructure (barriers to entry)
* Faster and more efficient research/data collection design design
* Reproducibility
* Credit for producing items in the life cycle
* Increased use of DDI; tool that’s used at (and enhances) all stages of research w/in infrastructure
* Purpose: support discovery, analysis, preservation, harmonization, reuse to enable future research more easily
* Inclusions and exclusions; can incorporate variety of observational data while intersecting with/relying up on other standards as appropriate
* Multi-lingual and multi-country environment
	1. What this document is not/caveats
1. Vision for long term infrastructure for the social sciences (George)
	1. Example use case: Survey design







1. Integrating DDI into the Data Lifecycle
	1. Element registry
	2. Instrument Design
	3. Data transformation
	4. Archiving
	5. Data Discovery
	6. Data extraction
	7. Data merge
	8. Data analysis
2. Lifecycles of other data types
	1. Administrative data (Katy?)
	2. Other data with existing metadata standards (e.g. EHR)
	3. Other object types: images, text, video, etc.
3. Stakeholders (Katy) (includes types (T), needs (N)
	1. Data producers:
		1. Research, administrative, other; academic, government, other
		2. T: Large-scale
			1. T: Survey designers
				1. T: Academic and government
				2. N: Re-use of existing survey components
				3. N: Design new survey components
				4. N: enable changes in measurement while maintaining comparability over time
				5. N: ability to integrate various types of measures (biometric, open-ended responses, etc.)
			2. T: Survey operations
				1. T: may/may not be one and the same as survey designers
				2. T: Academic, government, and commercial
				3. N: ease of receipt of requirements from designers
			3. N: Reproducibility/transparency (incl. ability to track their work)
			4. N: Greater efficiency in doing their research
			5. N: demonstrating use and impact of the data they produce
			6. N: Discoverability of their data
		3. T: Small-scale
			1. N: Ability to document their research/create metadata as they do their work w/minimal additional cost of using DDI (i.e., integrating into existing tools they use)
			2. N: Reproducibility/transparency (incl. ability to track their work)
			3. N: Greater efficiency in doing their research
			4. N: demonstrating use and impact of the data they produce
			5. N: Discoverability of their data
	2. Data subjects/survey participants
		1. N: confidentiality
		2. N: demonstration of public benefit resulting from their participation
	3. Data repositories
		1. T: broad, specialized, archives, self-publishing platforms
		2. N: provide access to data at various levels (given issues of licensing and confidentiality)
		3. N: Enable discovery and analysis by end-users; incl. enable discovery in tools beyond their own catalogs (e.g., virtual metadata pond, but also linking data and publications)
		4. N: Preservation (for some)
		5. N: Richer metadata accompanying deposits, as well as more to be added/tracked throughout their workflows (which will improve ease/speed of processing pipeline and provide a higher-quality product)
		6. N: demonstrating use and impact of the data they make available
		7. N: ability to build upon/learn from the work/systems/tools that others have done to better leverage resources
	4. Data users:: Secondary data users (note: includes academia and beyond (public sector, non- and for-profits, etc.) and people with various levels of skill and expertise
		1. N: Discovery (known-item and by features (topic, time, geography, etc.)) across multiple data sources and repositories; includes discovery at the variable level
		2. N: Analysis (remote and on desktop) (given varied skill levels)
		3. N: Linking and combining datasets from different sources (esp. in new ways)
		4. N: Comparing change (over time, geography, etc.)
		5. N: Reproducibility/transparency (incl. ability to track their work)
		6. N: Using data as part of teaching research methods
		7. [placeholder for referencing data producers as data users as well]
	5. Funding agencies
		1. Large-scale national agencies
			1. National Science Foundation (NSF)
			2. National Institutes of Health (NIH)
			3. Deutsche Forschungsgemeinschaft (DFG)
			4. Economic and Social Research Council (ESRC), and potentially others, within the broader umbrella for Research Councils (Department for Business, Energy and Industrial Strategy)
			5. Swiss National Science Foundation (SNF)
			6. Swiss Commission for Technology and Innovation (CTI)
			7. Fill in others here; especially international examples
		2. Large-scale international agencies
			1. European Commission (EC) - especially Horizon 2020 and Eurostars program
			2. Organisation for Economic Co-operation and Development (OECD)
			3. World Bank
		3. Ministries
		4. Private Research Foundations
			1. Alfred P. Sloan Foundation
			2. Wellcome Trust (esp. as relates to biomedical)
			3. Fill in others here
		5. Other Government
			1. Institute of Museum and Library Services
			2. Other
		6. Universities
	6. Other standards:
		1. N: having DDI complement their standard by documenting aspects they don’t cover
	7. (DDI) Tool/service developers (often would live in one of the aforementioned stakeholder organizations):
		1. N: Having their tools widely known and adopted
		2. N: Awareness of other tools upon which they can build
		3. N: Finding collaborators to co-create tools when a need is shared by multiple organizations
		4. N: Long-term hosting and maintenance of tools
	8. Members of the public:
		1. N: to gain a benefit for society from data gathered and used
	9. Other communities
4. Services and Tools (Ingo)
	1. Instrument design
	2. Data creation
	3. Documentation tools
	4. Need to define format for Datum Pond
	5. Need to define format for Metadata Pond (PIDs)
	6. Tools needed for Element Register
		1. Submission
		2. Curation
		3. Discovery
		4. Survey Design
5. Strategies for realizing the vision
	1. Building on existing tools and projects
	2. Utilizing related standards
	3. Research and demonstration projects
	4. How to engage stakeholders
6. Next steps
	1. Grant application components
	2. Publications
	3. Coordination