### Is this metadata management tool any use?

Extending CESSDA's software maturity matrix to the DDI domain



John Shepherdson Head of CESSDA Technical Work Group

# cessda

License: <u>CC BY 4.0</u> (for exceptions, see last slide)

### Structure

### CESSDA

- Overview, vision, objectives
- Common interoperability characteristics

### Ensuring software quality

- Technical framework
- Quality control and acceptance criteria
- Software Maturity Levels (SML) scoring

### Extending SML to DDI Tools

# cessda

CESSDA Technical Framework Workplan Task Deliverable D1: Technical Architecture (v1.0)

Authors		
Contributor	Service Provider	Contact Email
John Shepherdson	UK Data Service	john.shepherdson AT essex.ac.uk
Ørnulf Risnes	NSD	ornulf.risnes AT nsd.uib.no
Mike Priddy	DANS	mike.priddy AT dans.knaw.nl
Matti Heinonen	FSD	Matti.Heinonen AT staff.uta.fi
Johan Finn	SND	johan.finn AT snd.gu.se
Wolfgang Zenk-Möltgen	GESIS	wolfgang.zenk-moeltgen AT gesis.org





- Permanent legal entity owned and financed by the individual member states' ministry of research or a delegated institution
- Each member is represented by a national institution, a Service Provider
- Norway is hosting CESSDA, main office in Bergen
- Recognised as an ESFRI Landmark in the ESFRI 2016 Roadmap in the field of social and cultural innovation



### CESSDA's vision

The vision of CESSDA, as stated in its statutes, is to provide a **full scale sustainable research infrastructure** that enables the research community to conduct high-quality research which in turn leads to effective solutions to the major challenges facing society today



## CESSDA's objectives

- To support national and international research and cooperation in areas expected to be of great importance in the future
- To facilitate access to social science (and related areas) data resources for researchers regardless of the location of researcher or data
- To provide large scale, integrated and sustainable data services to the social sciences and facilitate and support research, teaching and learning

# CESSDA's 5 common interoperability characteristics

- **1. Loosely coupled but coordinated** enable Service Providers to retain independence, yet fully interact in an integrated service
- **1. Sustainable** enable medium and long term investment and business change decisions to be made



# CESSDA's 5 common interoperability characteristics

- **3. Extensible** enable additional services to be built on or around it, including adapting to changing functional requirements over time
- **3. Maintainable** enable components to be updated when IT specifications change
- **3. Standards based** enable the coordinated and planned change to all the coupled, but coordinated, services

### CESSDA's Technical Framework

A guide for the development of the various (software) products and services that form part of the CESSDA Research Infrastructure

- promote good practice for software development
- protect software assets
- meet common interoperability characteristics

Technical infrastructure for Development, Staging and Production

• harmonise software development tool chain for SPs

### Protection of Software Assets

Ensure CESSDA has access to

- source code
- configuration files
- technical documentation

For Research Infrastructure components



## Quality Control

Software Maturity Levels

- ensure quality of the research infrastructure is maintained
- guidance on minimum, expected and excellent standards
- originally based on <u>NASA's RRLs</u>
- revised in light of 'Capability Development Model' from <u>CESSDA SaW</u> project



### Maturity Modelling - More Info

See EDDI16 presentation:

<u>A Capability Development Model for Assessing and</u> <u>Improving Distributed Infrastructures and their Services</u>

Mike Priddy, Trond Kvamme, Marion Wittenberg



### 11 Product Acceptance Criteria

- Documentation
  - Development, Operational, End User

Ces

- Intellectual property issues
- Extensibility
- Modularity
- Packaging

# 11 Product Acceptance Criteria

- Portability
- Standards compliance
- Support
- Verification and testing
- Security
- Internationalisation and Localisation

Ces

### Software Maturity Levels - SML

0 - Not applicable

1 - Initial usability; software use is not recommended

**2 - Use is feasible;** the software can be used by skilled personnel but with considerable effort, cost and risk

**3 - Use is possible by most users;** with some effort, cost, and risk. A risk assessment should be made before use

4 - Software is usable; with little effort, cost, and risk

**5 - Demonstrable usability**; there is clear evidence that the software is widely used by many users

Cesso

### Intellectual Property

- 1. Software developers have been identified and their responsibilities have been determined.
- 2. Developer organisation(s) (or developers) have an agreement with CESSDA that addresses any potential conflicts in the proposed intellectual property rights and responsibilities for development.
- 3. Agreements on development responsibilities, the list of developers, a recommended citation, and intellectual property rights statements, offering limited rights for use, are available, perhaps upon request, for review.



### Intellectual Property

- 4. There is evidence that all developer organisation(s) (or developers) have confirmed that the list of developers, recommended citation, and intellectual property rights statements, including limited rights for use, in the software source code, documentation, and in the expression of the software upon execution, conform to CESSDA's policies and agreements.
  5. There is evidence that all developer organisation(s) (or
- developers) have confirmed that the list of developers,
  recommended citation, and intellectual property rights statements,
  including limited rights for use, in the software source code,
  documentation, and in the expression of the software upon
  execution, conform to CESSDA's policies and agreements.

### Intellectual Property

Developer sign up – get write access to CESSDA's code repos

Complete online contributor's agreement - re code ownership and IP

#### CESSDA Research Infrastructure Contributor License Agreement

CESSDA AS requires that You sign a Contributor License Agreement ("CLA") regarding any software code and/or documentation You wish to contribute to the CESSDA Research Infrastructure ("Contribution").

By submitting your Contribution to the CESSDA Research Infrastructure Project, You hereby agree to license your Contribution under the Apache License, Version 2.0 (https://www.apache.org/licenses/LICENSE-2.0), and to include the appropriate copyright notice required by the license.

Cess

#### \*Required

Given name(s) \* e.g. Jane

Your answer

Family name \* e.g. Smith

Your answer

Email address \*

Your answer

#### Bitbucket account name \*

Your answer

## Software Maturity Levels Matrix

Intellectual Property (CA2)	Celevability (CA3)	Modularity (2.44)	Pedraging (CAS)	Autability (CA6)	Standards Compliance (CAP)	Support (CAB)	Welfication and Yesting (CAO)	Security (CADI)	Internationalisation and Localization (CALL)
Software developers have been elemented and their responsibilities have been determined. Referent policies of developers of gammafactual for developers (have been entermented for applicability for interferciant property policies. These employ endoes of a doubt interferciant property rights agreement that would recall from supportable activities with other developer agrehadings in developers, Bajtis are not possified themeship for the survey calls at referentially in decommentation, are rights are limitations have not have specifical.	The software was not despised with softwarehold or read, so there is within an addity is indeed or modify program believes, or it is easy difficult to do, even for angue, similar to these of the software sharpest. There is no, or limited, and addity of the snare society for lenging the software scales to follow, with limits to no software.	These a evidence that the source code was written with no design or cannot design in a source of the source of the functionality for moduliarity or use. It may have been a demonstration or pflat project.	Only powers code or executable swelleds (i.e. on packaging there is no or incomplete, including the documentation and no auto-build/installation facility is analative. Item an experienced ator may have difficulties installing the software.	The cobserve to a whole a not portable, however, if there is surface source code provided with sufficient interval, endered and contervised with sufficient interval, endered and contervised in any large particular. Towardshife histories are possible for a specific platform, has there is no surface information any port- late there is no surface information any port- porting as a whole is not funchly (a g, due to licensing) or prohibitisely expension.	The software and adheses development across compli- ational to part, with loading defined databased and hand produces. The clandide's may be interesting, are externally deverting, have a may be implemented with modifications to meet load conditions. These may be infer are not construct endowce of clandiards word, but it implempending in the influences.	There is known-contact information avoidable that the developer organisation (s) and there is a willingeness for provide minimal, samanismot support attilisant guarantiers. It may not be parallele for an analuser to say the samalife for an analuser to say the saffware without some support.	Software suplication form ultitod and unit facility performed.	becardy was addressed in the development planes or to and including design.	Menophaneliation and Localization not addressed
Developer organisational (or developer) have an apreement that addresses are potational conflicts in the proposed interfactual property rights and respondetities for development. A finited right, statement has been during, and applied incombinity in documentations and waves ranks. Breatloper organization(o) (or directoperc) may be serial and is regulation rights for seri-	Where is some consideration to -semanbing, but that may only used. For a human number of use cases, through use of methods such as adjust-administed darks on athem stands which provide legals durbnics. However estimatelity is provide transplay durbnics, there are administrative in an adjustmition providers of assume the application clearly identified services of source code; and/or limited apportunity for software modification.	These is to-distinction between generic and substan-specific functionality. The source cade is organized into a personny option that provide general functionality and area or two indepthene that south provide multiple, semilated, functions, used within south multiple and south multiple, densitient and an extensibility of the multiple and south multiple, and area or two indepthene is closed with only a few interval functions according by external programs through the primary spitem.	Software includes anti-level features, har a only exercise operating system. Detailed includion including an available and hundling for other operating systems is possible for an importement asis.	The complete source code is excluded, without enterind dependencies that are not portable, but the software cannot be jorital without significant damps, is the suffware the large canonics. Sourcementation on parting is incufficient, and should be taken bits compared to reacting for the near context respland. Cost-benefit and the malyzer, compared to reacting for the near context and use cases, should be undertaken before porting is considered, and only instant of the case benefits of using the collever sightly suffward; then cast of developing new suffwares. Nutling will memblishers require suffwares. Nutling will memblishers require suffware filed.	The applears and atheres development process endoercor to comply with endoercor to comply with endoercor to comply with many and the proclement has authors are confluction or transfer of any and to complete. These maple some documental evidence that may not be complete.	The developer experimention(2) sequent to reported masses with updates/paticles; that are usually made available in a massaulty linearly tachies. Items sequent is available, out may be interestimetation. There is within an an allogeneral ware community that provides arousers, for example, via a mailing list or holistic bood. Documentation and source code availability may be autificated for an experimentation and an orace of an endrouser list not require extension or and-user/1s not require extension tappent.	Software supplication doministrated and tracted to a full-software context. Nating includes feating for error wonditions, and proof of handling of undersome topol.	Society was addressed in the development phones up to and including implementation. Developms, have andertaken appropriate Texacilip tasking.	Software is locale aware
Appenentry on development reagonalitiking, the lot of developers, a recommended contrain- and introduction properly right automonity. Others for the lot of developers and the performance of the second second second second second second second second second second second the second	Evolution excernationally a designed into the automn for an orderind excerning of one cause. The procedures for antioniding the culturary are dufined, whether by same used modification or through the provides of arms type of withorhan framework including to g-, culturate basels on surging supatilities, it where survex code modification is possibilities, if where survex code modification is possibilities, it where survex code modification is possible of the orientee plan, the software is wall or unstand, where a modernate no high level of orderation, and has a configuration elements dearly separated from logic and display elements.	Phene is evidence that the architecture is open, with full discultance provide velocitation prevents that provide from Dane or previous the autoxide wolf-ther. E.e., open architecture), internal familians an annukan beginnen full familians, that much hereitige resultion have here so-resoluted for generalized familians that much hereitige have manualed the added from specified from families have much have manualed for all of the specified from families, such within much much added contains, many independent legical paths.	The software is easily configurable for diffuent contract such as buckness of measures (the investment, STM) are unifyerable. All configuration specific information is contratised.	The otherway is moderately portable. The sufficiency on the portable of the only institutely sufficiency and the portable of the only institutely sufficiency fixed. Decommodations are porting which and its sampletine, taken requires considerationality collect and segmethics. Know realistication collect and segmethics. Know realistication collection and segmethics and the sampletine set of the conderlying sufficience on the target spream may be necessary.	We addresses and addresses development process comply with speer, encaptiond or proprietary damkards, hal- there is insuranghine compliances in encaptional translation of compliances. Compliances in encaption damkards has be rested but visio may raw be for all components. There is documents who allows of documents who allows of documents of allows of documents.	Support is controlled in a website confarming evolves of moursel, accesses to MRL other useful referenation and a sammarily support agentiles. If ansare income long, building or mercupy/blaunable income long and and an ansare income and an ansare and an analysis engages with users in the community support seas. There are suppler transtatived releases of updates/patches that are made available and upport for a support seas. Therein an suppler language at the location of updates/ patches the location of updates/ analysis and an analysis of updates/ analysis and an analysis of updates/ analysis of the location of updates/ hereins formal Agenerates.	Software supplication downsmithed, minimal and validated in a redenant Contant.	Security was addressed in the development dramer up to and including implementation.	Content to localize is enternal to-more cade
There is evidence that all developer organizationfol for developerd have confirmed that the full of developers, recommended ultation, and intellectual property rights attaineeds, including limited rights for users in the reflexers source code, documentation, and in the representation of the sufficience space encoded limit their insidiational publics, and apprements. There includes are player encoded limit to limit and payment of and guardines or their insidiation and the sufficience space encoded limit to the sufficience of the representations, machine resultable code supervising intellimities guardines are their representations, machine resultable code supervising intellimities (and consiste dataments in language that is no to anderstand by fuggersons, such as a per services, and conditions for use. (New Instances on experiments) in face spaces, excitations, and conditions for use. (See all statements on experimental indications) (and any space instance of single point to obtain formal statements) on statements of rights or to negative additional rights.	The sectorability capability for the software is well defined, based sampled our case, providing many points of vestimistify. A distinct enternative plan is publically evaluated enternative plan is guidencost developers to income familie wells' project to entered the software is a reasonable ansared of the software is a software of the project of the software is a function of a software optimistic factors on a separation. These is evaluate the software has been extended and applied to a contex of a software factor. For it we project, wing optimistic documentation, but may have breaked on the and software factor. But may have breaked and how well software factors.	There is clear organisation of all components into Meanies or on rate registines with constituent documentation of all Meanies and PATs or standard end on row row the Tacas. Modelin have been orwarded for all uport/Path Tara lians and upperhad his filter is with eventher for functions within interfaces. Tail second his filter is with mailede may contrain many independent legical paths.	On installation there is this-batter; and auto-baild for incore than one supported platters with 28 defection configuration files autifield entry with solid of the generative field to plattere-respectly. texting, United allation is researched in sightfilement with the penalefility is remove all control files, and configurations.	The software is highly portable. The software can be parted to all find the most obscure or obsolve softwares without worthcallers. The documentation is complete and thereagh. Be documentation is unsplicit and the software here affect to part the software is minimal.	The software and software development process comply with recognized and proprietary standards. Compliance with Home standards has been writhed through tracing for all components. Incommende auktores for solicetail standards and the use/forthet through teating is available.	Here is separated and clearly-defined rapport by the developer with an email heightest and additional documentation heightest and additional documentation federmatilies for a compro of our externation (developer), operations staff, and endiouxed). There is explain culdence there are constrainty of support is implied in maybe possible to support is implied in maybe possible to support a standard offering of the developer organizationial.	Actual software application "youtherd" through text. and demonstration itemats requiremented and suscendually definenced.	Security eas addressed in the development phones up to and including verification and tecting.	Contrast to localize frost, lepost, apaphica and multimedia. Network, bande data and sharasilee with, bande pasaren   has konen internationalized.
New one multiple attacheds withoutlood into the sufficience product delaciting wheeth could rights and any conditions for wax, including commercial and new-commercial wax, and the	where is an elements that the settingent has been instanded externally by scent actuals of the original	It is exclude that of functions and data are an appalated into objects or accessible through well-service interfaces. There is	A user interface public the inclution to sugh a findage needed to build, cardigare, and includ the call even. An exercisitient sature	the software a completate pertains. The call arene can be perted to all cellance clear it	Compliance with open or internationality receptions	No support by the organization of a clearly defined with thequest and timely	Actual activate spatication feated and willcated through caccersful are of	Secondly was addressed in the development phases up to and	Software has been tested with provide translations.
recore resented clathion. The field of devide, peri in endended in five sociate cardioof the periods, is the documentation and in the sequencies of the socializer opprecisation. The indefinition property rights chalorevers are emprecade in logalizergange, reaching-reaching- ender, and in sometic statements in long ange that out he order trand by legentary, such as a per written, reacquitable forms.	devolutions of a row using exciting the activitation and "Term is a low approach the modifying and nativeling fractions in source and the modifying and intervaling fractions in an activity of the modifying of the holding of a devolutions and the area and access a range of classical in providing a consignment operand for accession and over generated connect for accessions and over generated documentation an extransion is also available.	constructed on two Teac-Teag will's instancinghal (instructing the attribu- card use or all generative instances have been specific to a super-teaching and competibilities (instances) of the specific teaching and complexities (instances) devices and analytic attribute constraints. Instances and analytic attribute constraints, lieu insideported and logical paths.		rom, en un augella diros lapor a direc. Elano an Elan anderidopia quen alleg quenes lapor lapor facilitariane la architecto la languagon direce, Cili, edu- lacideneng al liandi, directarilhanare e el la socia an any optimiento architecto directargenego inte- angolicatelan laguer hao toxen-instrudiod.	classifields for the schware and sufface demolgrammi- process, is estimated and documented, and estified through testing of all comparison without and hospital and the state of the documented through signaler stating and cartification from an independent process.	problem, informers, util, surger utility for the serveds of Dire one vacuum tillings, as well as cannabilitation of champes by the cannonicity. There is a chaffed independent brighted as a validited in independent of the service of the signaling proops an activities and independent exploite surger fram the developential and developer oppresentations. There is mathewer that candinality of surgers is implementation of the surgers in implementation of the surgers in the based of a surgers in the three of the based of a surgers in the three of the based of a surgers in the developential cand dig surveyers (EDA) with the developent(C) as		ndering product where	

Cessc

6

Colour	Meaning
	Minimum standar
	Expected standar
	Excellent standar

### Online Form

### **CESSDA Software Maturity Levels**

Form v02.00, 23 September 2016

\*Required

#### Using this form

You can use this form to assess a product (typically a 3rd party product you are thinking of adopting, or a component that is being considered for use as part of the CESSDA Research Infrastructure).

The information you enter will be stored in CESSDA's Software Maturity Levels (SML) scorecard. When you reach the end of the form, you can choose to have the scores you entered emailed to you, along with the overall product assessment.

There are eleven criteria for you to score against. Each can be scored from Level 1 (low) to 5 (high), or 0 (not applicable).

Each of the criteria has a minimum standard associated with it, which should be met or exceeded by any component intended for use as part of the CESSDA Research Infrastructure.

#### Product name \*

Your answer

### https://goo.gl/forms/uwuye0nTUkti7AiH2

Cesso

## Confirmation Email

Thanks for using the CESSDA Software Maturity Levels (SML) form.

#### The overall score for Open Source Metadata Harvester (OSMH) is Level 3

Which means "Software is usable; the software can be used by most users although there may be some cost and risk" (i.e. expected standard for CESSDA RI use).

You entered the following values:

CA1.1: End user Documentation	0
CA1.2: Operational documentation	2
CA1.3: Development Documentation	2
CA2: Intellectual Property	3
CA3: Extensibility	4
CA4: Modularity	3
CA5: Packaging	2
CA6: Portability	4
CA7: Standards Compliance	2
CA8: Support	2
CA9: Verification and Testing	3
CA10: Security	3
CA11: Internationalisation and Localisation	3

and provided the following feedback:

"It would sometimes be good to be able to comment on a score, because there are nuances and trade-offs here. Also, an N/A-category could be useful. Used 0 for the CA1.1 since the OSMH don't have "end users" as such."

Regards,

The CESSDA Technical Work Group

Not applicable

Ces

# Extending SML for DDI Tools

### Some suggested criteria:

- Imports/exports multiple versions of DDI
- Multiple representation formats supported (formal syntax)
- Has declared semantics (uses CVs, thesauri, ontologies ...)
- Maintains Provenance
- Supports Curation
- Data cleansing, consistency checking
- etc.

### Extending SML for DDI Tools

Straw poll

Is this approach any use?

Please raise an arm if you think so



### Extending SML for DDI Tools

Why useful?

Could reduce barriers to acceptance and reuse

- DDI tools
- DDI metadata

More discussion at panel session 'Re-Use of Software and Administered Metadata' on Wednesday at 15:30

### Conclusion

Software reuse is by design, not by accident.

Adding 'reusability' at the end is time consuming and expensive



# Thanks for listening

Any question?





# Cessda

—in your circle of trust.

*website:* <u>www.cessda.net</u> / *twitter:* @CESSDA\_Data

### License

This presentation is offered under license <u>CC-BY 4.0</u>

The license does not apply to the following copyrighted material used in this presentation:

• Photograph of the CESSDA Main Office



• CESSDA logotype



• CESSDA colour bar (present at bottom of this and every slide)

