

How  
to

**share**

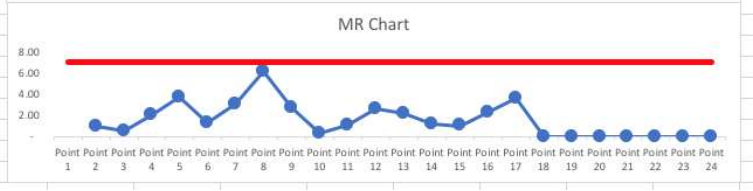
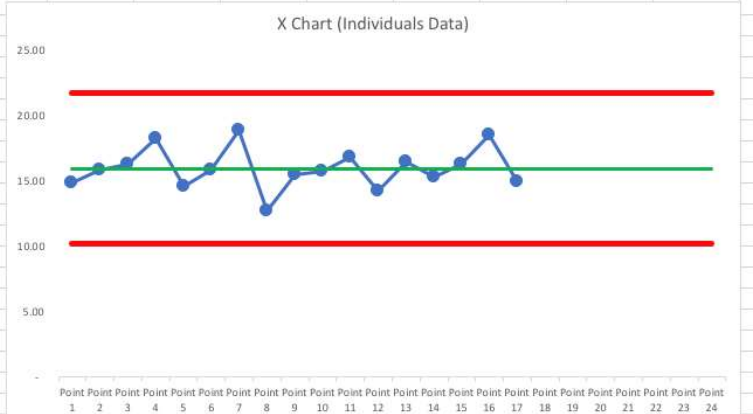
Language data



What is language data?

What we think as (language) data

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T		
1	<b>Process Behavior Template provided by Mark Graban, author of "Measures of Success"</b>																					
2		<b>Data</b>	<b>Average</b>	<b>Moving Range (MR)</b>	<b>MR Bar</b>	<b>Lower Natural Process Limit</b>	<b>Upper Natural Process Limit</b>	<b>MR Upper Range Limit</b>	<b>Target</b>	<b>Halfway to Lower Limit</b>	<b>Halfway to Upper Limit</b>											
3	Point 1	14.89	15.98		2.17	10.22	21.75	7.08		13.10	10.87											
4	Point 2	15.83	15.98	0.94	2.17	10.22	21.75	7.08		13.10	10.87											
5	Point 3	16.32	15.98	0.49	2.17	10.22	21.75	7.08		13.10	10.87											
6	Point 4	18.32	15.98	2.00	2.17	10.22	21.75	7.08		13.10	10.87											
7	Point 5	14.58	15.98	3.74	2.17	10.22	21.75	7.08		13.10	10.87											
8	Point 6	15.90	15.98	1.32	2.17	10.22	21.75	7.08		13.10	10.87											
9	Point 7	18.90	15.98	3.00	2.17	10.22	21.75	7.08		13.10	10.87											
10	Point 8	12.76	15.98	6.14	2.17	10.22	21.75	7.08		13.10	10.87											
11	Point 9	15.50	15.98	2.74	2.17	10.22	21.75	7.08		13.10	10.87											
12	Point 10	15.80	15.98	0.30	2.17	10.22	21.75	7.08		13.10	10.87											
13	Point 11	16.90	15.98	1.10	2.17	10.22	21.75	7.08		13.10	10.87											
14	Point 12	14.30	15.98	2.60	2.17	10.22	21.75	7.08		13.10	10.87											
15	Point 13	16.50	15.98	2.20	2.17	10.22	21.75	7.08		13.10	10.87											
16	Point 14	15.30	15.98	1.20	2.17	10.22	21.75	7.08		13.10	10.87											
17	Point 15	16.30	15.98	1.00	2.17	10.22	21.75	7.08		13.10	10.87											
18	Point 16	18.60	15.98	2.30	2.17	10.22	21.75	7.08		13.10	10.87											
19	Point 17	15.00	15.98	3.60	2.17	10.22	21.75	7.08		13.10	10.87											
20	Point 18		15.98		2.17	10.22	21.75	7.08		13.10	10.87											
21	Point 19		15.98		2.17	10.22	21.75	7.08		13.10	10.87											
22	Point 20		15.98		2.17	10.22	21.75	7.08		13.10	10.87											
23	Point 21		15.98		2.17	10.22	21.75	7.08		13.10	10.87											
24	Point 22		15.98		2.17	10.22	21.75	7.08		13.10	10.87											
25	Point 23		15.98		2.17	10.22	21.75	7.08		13.10	10.87											
26	Point 24		15.98		2.17	10.22	21.75	7.08		13.10	10.87											
27	Point 25		15.98		2.17	10.22	21.75	7.08		13.10	10.87											
28	Point 26		15.98		2.17	10.22	21.75	7.08		13.10	10.87											
29	Point 27		15.98		2.17	10.22	21.75	7.08		13.10	10.87											
30	Point 28		15.98		2.17	10.22	21.75	7.08		13.10	10.87											
31	Point 29		15.98		2.17	10.22	21.75	7.08		13.10	10.87											



What language data actually is

## Dramatis Personæ.

Julius Cæsar	Mr. Gadsden
Q. P. Cato	Mr. Fry
Antony	Mr. K.rough
Brutus	Mr. Bannister
Cassius	Mr. Seal
Calpurnia	Mr. Graft
Portia	Mr. Swindon
Decius Brutus	Mr. Swann
Marcus Cicerone	Mr. Williams
Cinna	Mr. Harcourt
Artimedes	Mr. Gable
Messala	Mr. Fenner
Titinius	Mr. Walford
Cinna the Poet	and
Flavius	Mr. Goss
	Mr. Jones
	Mr. Smith
	Mr. Davenport
	Mr. Lee
	Mr. Blyth

Women.

Calpurnia                      Misses Dugdale  
Portia                         Mrs. Cook

Guards and Attendants.

Scene *ROME.*

THE

## THE TRAGEDY. OF JULIUS CÆSAR.

*After Titinius. Scene Fifth.*

*Enter Titinius, Calpurnia, and several Attendants upon the Stage.*

*Titinius.*  
**H**ow! how you tell Calpurnia, you poor women  
I die a Soldier? What, know you not  
I'm a Soldier? you might not walk  
Down a Whorehouse, without the light  
Of your Husband's Spurs, when Truce is over?  
*Calpurnia.* Who he, a Soldier?  
*Titinius.* Who he, the Leader down, and the Eagle?  
Who doth down with the best Apparel?  
You he, when Truce is over?  
*Calpurnia.* O! Titinius, in respect of a few Workmen, I will bet at your  
word by a Gallus.  
*Titinius.* O! what Truce are these? Against me directly.  
*Calpurnia.* A Truce he, that I hope I may see, with a Gallus.  
Which is indeed by a Master of his Trade.  
*Titinius.* O! what Truce are these? Thanksgiving hours, what Truce?  
*Calpurnia.* O! Titinius









Why, and if we can't get that some information and see if we can help. What is your first name?

**Fred**

Fred.

**Emily**

And your last name; would you spell it for me please? Fred Sure. It's C-H-A-M-B-E-R-S, Chambers.

**Emily**

Okay. And your company name?

**Fred**

I'm with GoldStar Environmental.

**Emily**

GoldStar Environmental?

**Fred**

Yes ma'am.

**Emily**

Okay, and what's your phone number?

System:

**Fred**

No, actually, I can't even get an Internet connection.

**Emily**

Okay. According to my records, the AmStar700 is a voiceover Internet protocol phone. It appears that because your Internet isn't working, your phones are not working as well. Do you know who your Internet provider is?

**Fred**

I have Verizon.

**Emily**

Okay. Mr. Chambers, I'm going to get a hold of Randy, and have him return your call. Is the 610-265-1715 a good number to reach you at right now?

**Fred**

Yeah that's my cell. That is working.







Legal aspects

Copyright

Sui Generis Database Right

Personal Data Protection

Free flow of non-personal data

Public Sector Information

Special Regimes (INSPIRE/ Cultural Data/ Open Science)

Confidential Information/ Trade Secrets

## Language Data:

- May be copyrighted Works
- Sui Generis Data bases
- Data that need to flow between countries
- May contain confidential information/ trade secrets
- May contain personal data
- May contain geodata
- May contain/ be public sector information
- May be protected as cultural heritage

# Regimes of Property

- Copyright I
- Sui Generis Data Base Right (SGDB)



# Regimes of Protection

- Personal Data (GDPR)
- Confidential Information/ Trade Secrets
- Cultural Heritage
- Research Ethics

# Regimes of Openness

- PSI / Open Data
- Open Culture (Europeana)
- Open Science
- Free flow of non-personal information
- Copyright II
  - Public Domain
  - Limitations and Exceptions/ Fair Use/ Fair Dealing
  - Open Licences
- Patent publications and metadata

# 3

Considerations and Questions

Rights

**IN**

**≥**

Rights

**OUT**

# Who am I?

**Producer/ Collector** of  
Language Data

> Need to obtain all  
necessary rights

**Disseminator** of  
Language Data

> Need to use the  
licences that support my  
value model

# 1. Is it mine?



- 3<sup>rd</sup> Party Copyright/ SGDB
- Personal Data
- Confidential Information
  - Trade secrets
  - Confidential
- Cultural Heritage Restrictions

## 2. Is it truly mine?

### A. Copyright

1. Public Domain
2. Am I covered by a limitation?
3. Is the correct licence there?
  1. Open Licence
  2. Compatible Open Licence

C. Is it clear of any **confidential** information?

### B. GDPR

1. Scientific Research Exception
2. Public Interest/ Mission
3. Legal Obligation
4. (Legal Interest)
5. Prior Informed Consent (PIC)

D. Have I obtained all **cultural heritage** licences?

### 3. How to share I (housekeeping)

---

- Record data provenance
- Record ownership (owner/ type of rights)
- Record limitations (e.g. classified/ confidential)a
- Record personal data and their legal basis
- Record licensing information
- Include notice and takedown procedure





## 4. How to Share II (licensing)

- Include a PSI notice if applicable in your jurisdiction
  - Notice new PSI rules!
- Use standard licences (Creative Commons)
- Use Free Cultural Heritage Works Licences
  - CC Zero
  - CC BY 4.0 int.
  - CC BY SA 4.0 int
- If you use additional restrictions (ND/ NC)
  - Always provide a way to obtain more rights (through additional licences)
- If you use custom licences:
  - Understand what your value model is
  - Ensure no other standard licences exist
  - Ensure they do not contain any copyleft elements
  - Ensure they are machine readable
- If you use restrictive licences:
  - Ensure you reduce frictions through electronic transactions

## 5. How to Share III (infrastructure)

- Use standard Free/ Open Source Technologies for the repositories (e.g. CKAN/ Dspace etc)
- Use the appropriate Open Data Standards (data/ metadata/ Services)
- Use (open) APIs
- Identify core data sets (see new PSI directive)
- Interoperate with major initiatives (ELRC/ EOSC/ OpenAIRE/ Europeana/ Wikimedia)
- Follow FA(I)R principles
  - Findable
  - Accessible
  - Interoperable
  - Reusable



**4**

conclusions

# Conclusions

- Language data are (mostly) not CSV files
- There are several legal regimes covering Language Data (some permissive – some restrictive) – we need to make the best use of them
- Always clear your content/ language data before your release
- Clearing is not just about copyright
- Watch the new PSI implementations
- Follow the Keep It Stupid Simple (KISS) principle in all licensing choices
- Use existing open infrastructure
- Interoperate with major initiatives – national, international, European, public and private

