# Steps and criteria for quality assessment of geoscientific data: from exploration to mine

**Yuri Parra Castilho**, Bruno Nunes, Gláucia Cuchierato

GeoAnsata Projetos e Serviços em Geologia

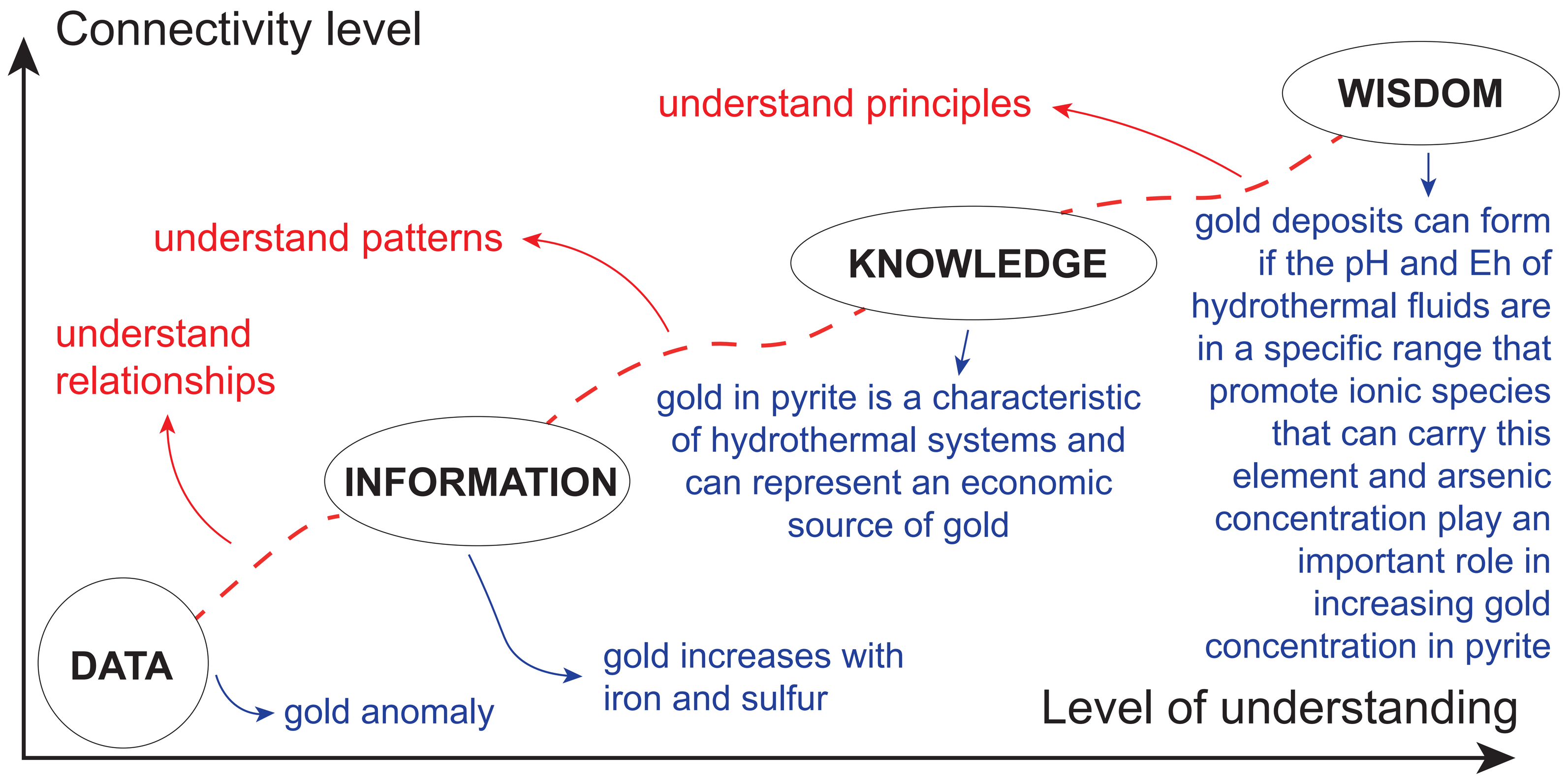
*yuri.castilho@geoansata.com.br (Castilho, Y.P.)*
*glaucia@geoansata.com.br (Cuchierato, G.)*
*bruno.nunes@geoansata.com.br (Nunes, B.)*

X SIMPÓSIO BRASILEIRO DE EXPLORAÇÃO MINERAL / X BRAZILIAN SYMPOSIUM ON MINERAL EXPLORATION

GeoAnsata — Transformando Recursos em Riquezas

## Introduction

- Around 2.5x10E+18 bytes (1 milion terabytes) of data is generated every day [1]
- Information implies in the processing of data to gain utility and answer questions such as "what", "who", "when" and "where", while knowledge answer "how" [2]
- The duty of big data is to unravel useful correlations for governments, companies and scientists [3]
- The "Big Earth Data" represents a frontier for geoscientists and machine learning tools need to be carefully developed to extract meaningful information [4, 5]




## Materials and Methods

Principles of Information Systems [6]

↓

Guidelines of international codes for reporting of mineral resources and mineral reserves [7]

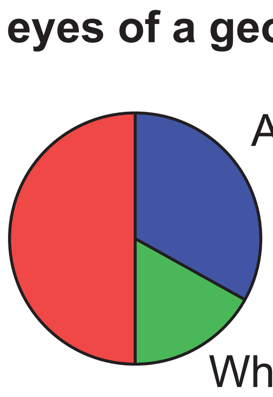
↓

GeoData Quality Management® [8]

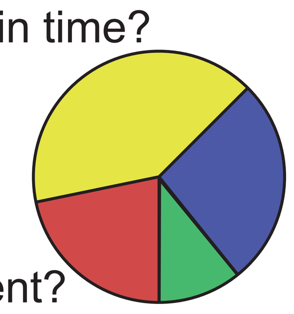
## Results

### The same database can offer different information depending on the experience of the observer:

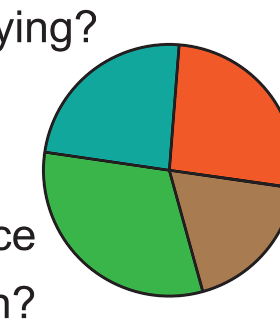
**The eyes of a geologist usually see:**




**Add a little knowledge on data science:**

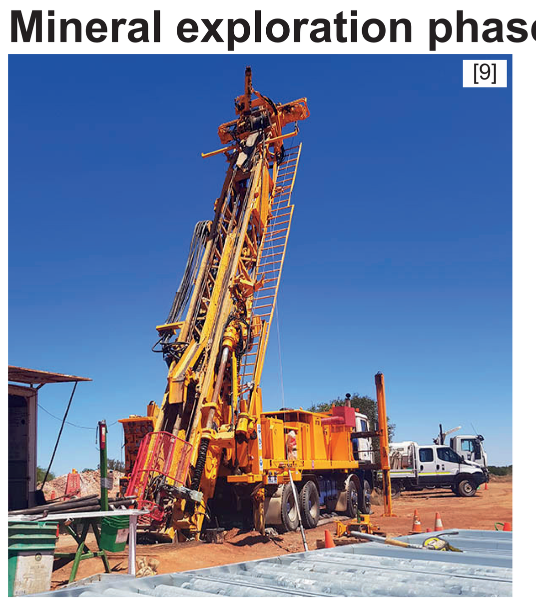



**Add a little of knowledge on CRIRSCO standards:**




### General features and sources of problems in geoscientific databases:

**Mineral exploration phase, lasting up to 1 decade**

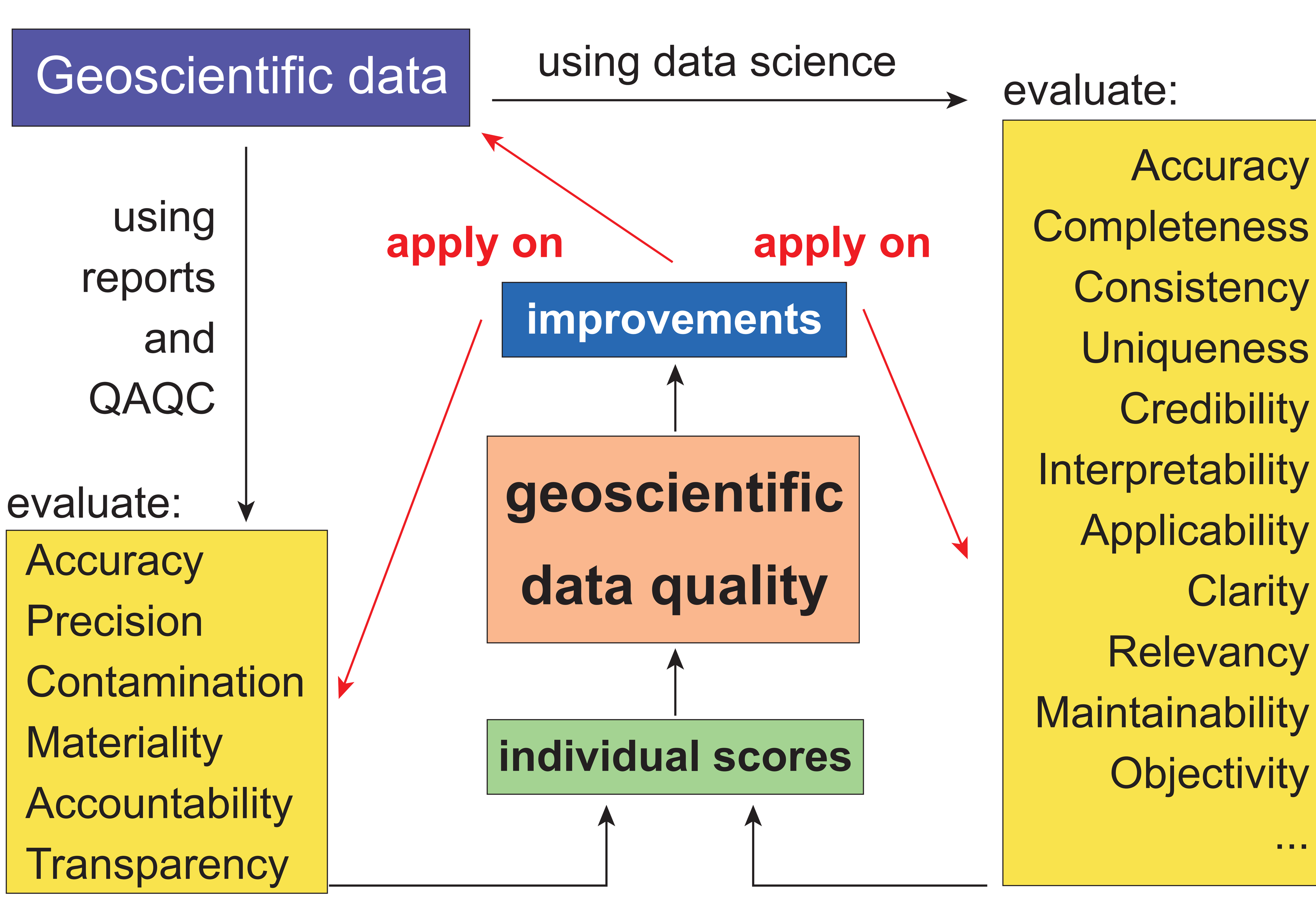



| General features | Consequences |
|---|---|
| Small data production rate (thousands of lines / year) | Lower concern about data quality but problems are easier and cheaper to fix |
| Low knowledge of the deposit or mineral system | Lower number of variables collected, lower the capability to recognize errors or outliers |
| High risk, high expectations, high pressure for mineralized drillholes | Preoccupation with showing results generates misinterpretations of the mineral system |
| Smaller technical team, easy communication | Managers may not agree upon what the field team defined, promoting rework in the database |
| Main areas: geology, geophysics, geochemistry | The database may not be structured or correlated, but the negative impact is lower |

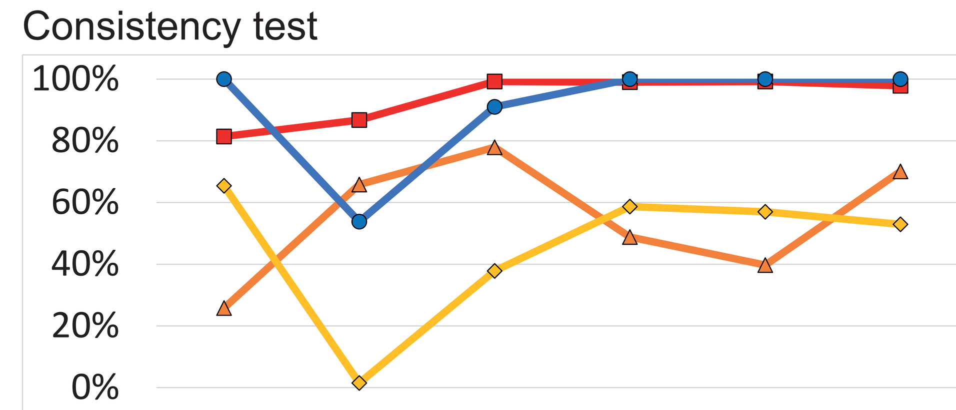
**Mine (extraction) phase, variable duration time from less than a decade up to several decades**

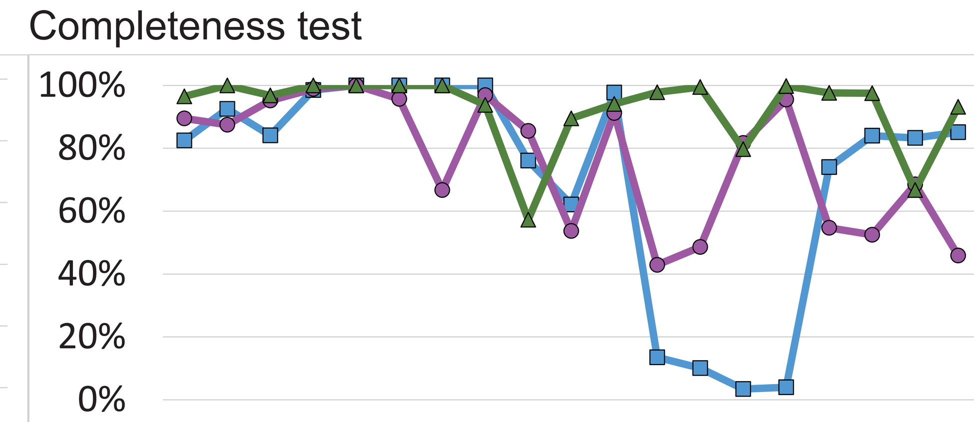

| General features | Consequences |
|---|---|
| Fast data production rate (thousands of lines / month) | Higher concern about data quality, but problems are harder and more expensive to fix |
| High knowledge of the deposit or mineral system | Higher number of variables collected, not enough time to recognize errors or outliers |
| Lower risk, high expectations, high pressure for consistent and continuous production | Pressure for production induces sloppy procedures for data collection |
| Larger technical team, difficult communication | Managers from different areas may not communicate correctly, inducing inconsistency, vacancy and rework |
| Main areas: geology, geochemistry, hydrogeology, geotechnics, geometallurgy | If the database is not structured or correlated, the negative impact is huge |

### Hypothetical flowchart for data quality assessment:




### Evolution of criteria used to assess data quality:




## Conclusion

Data quality is an abstract, ever-changing concept that depends on the commodity and the people involved in the activity.

It can measure maturity, accountability, transparency, competence, materiality while being used to improve techniques and procedures reducing costs and risks for mining activities.

## References

[1] htztps://www.forbes.com/sites/bernardmarr/2018/05/21/how-much -data-do-we-create-every-day-the-mind-blowing-stats-everyone -should-read/?sh=2774bf2360ba.

[2] Chen, M., Ebert, D., Laramee, R.S., Van Liere, R., Ma, K., Ribarsky, W., Scheuermann, G., Silver, D. (2009). ***Data, Information, and Knowledge in Visualization***, IEEE Computer Graphics and Applications, 29(1), 12-19.

[3] Sagiroglu, S., Sinanc, D. (2013). ***Big data: A review***, 2013 International Conference on Collaboration Technologies and Systems (CTS), 42-47.

[4] Guo, H. (2017). ***Big Earth data: A new frontier in Earth and information sciences***, Big Earth Data, 1(1-2), 4-20.

[5] Karpatne, A., Ebert-Uphoff, I., Ravela, S., Babaie, H.A., Kumar, V. (2019). ***Machine Learning for the Geosciences: Challenges and Opportunities***, IEEE Transactions on Knowledge and Data Engineering, 31(8), 1544-1554.

[6] DeLone, W.H., McLean, E.R. (1992). ***Information System Success: the quest for the independent variables***, Information Systems Research, 3(1), 60-95.

[7] Batini, C., Cappiello, C., Francalanci, C., Maurino, A. (2009). ***Methodologies for data quality assessment and improvement***, ACM Computing Surveys (CSUR), 41(3), 1-52.

[8] Cuchierato, Gláucia. ***A importância da qualidade da informação no processo de declaração de recursos minerais***. (2022). Tese de Doutorado - Escola Politécnica, Universidade de São Paulo, São Paulo.

[9] https://www.westcore.com.au/drilling_equipment_UDR1200HC.html [10] https://www.azomining.com/Article.aspx?ArticleID=1482