

Metadata for Beginners

Dr. Özlem ÖZKAN

Helmholtz Metadata Collaboration (HMC) Hub Matter



Who am I?



Dr. Özlem Özkan

BSc: Computer Education

MSc & PhD: Medical Informatics

Experience:

- METU Research Assistant,
- KPMG Data Scientist,
- MDC Research Data Manager,
- HMC Data Policy Expert &
- Training Officer



Gerlich, Silke. C., Strupp, A., Hofmann, V., & Sandfeld, S. (2023). Fundamentals of Scientific Metadata (1.0.0). Zenodo.

<https://doi.org/10.5281/zenodo.10091847>

Helmholtz Metadata Collaboration - Mission

Making Helmholtz data treasures visible!

- help researchers describe their data with **high-quality metadata**
- enable researchers to **reuse** of Helmholtz research data
- provide **advice, information & tools**

There are 6 domain specific hubs for each research field:

- **Matter**
- **Health**
- **Energy**
- **Aeronautics, Space and Transport**
- **Earth and Environment**
- **Information**



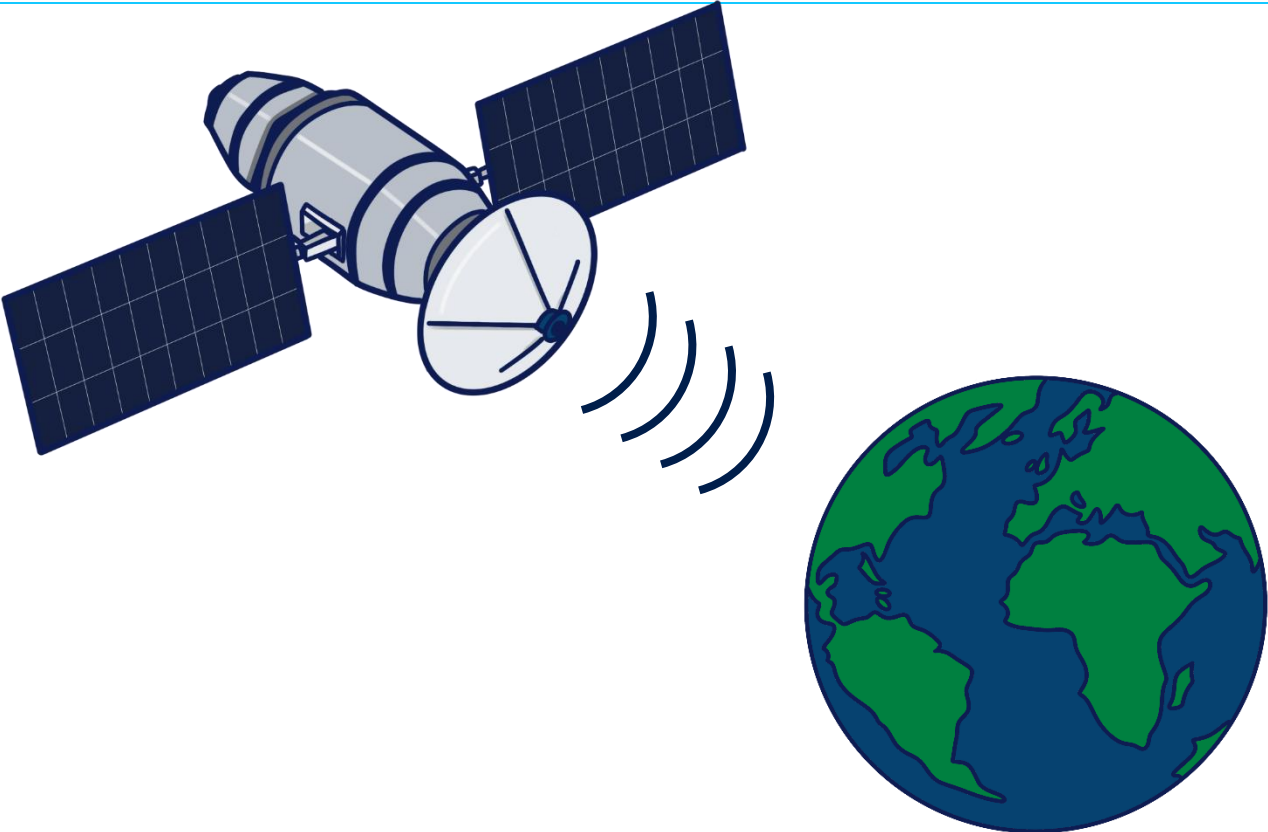
What can you say about this data?

	A	B	C	D	E
1	t	ax	ay	az	scr
2	0	0.3931848	-0.1593144	-0.4178079	0
3	0.01	0.3957354	-0.15696	-0.4242825	0
4	0.04	0.4138839	-0.1547037	-0.429678	0
5	0.05	0.4415481	-0.1512702	-0.4325229	0
6	0.06	0.4741173	-0.1488177	-0.434583	0
7	0.08	0.5021739	-0.1521531	-0.4285008	0
8	0.1	0.5247369	-0.1669662	-0.420849	0
9	0.11	0.5421987	-0.1813869	-0.4160421	0
10	0.14	0.5506353	-0.1947285	-0.4094694	0
11	0.15	0.5538726	-0.203067	-0.4057416	0
12	0.16	0.5534802	-0.2035575	-0.4056435	0
13	0.17	0.5527935	-0.1961019	-0.4098618	0
14	0.2	0.558189	-0.1908045	-0.4121181	0
15	0.21	0.5764356	-0.1865862	-0.4162383	0
16	0.22	0.589581	-0.18639	-0.4258521	0
17	0.25	0.6049827	-0.1941399	-0.4243806	0
18	0.26	0.619992	-0.206991	-0.4192794	0
19	0.27	0.6320583	-0.2191554	-0.4092732	0
20	0.3	0.6392196	-0.2279844	-0.3975993	0
21	0.31	0.6465771	-0.2317122	-0.3908304	0
22	0.32	0.6583491	-0.2291616	-0.3950487	0
23	0.34	0.6725736	-0.2220984	-0.4050549	0

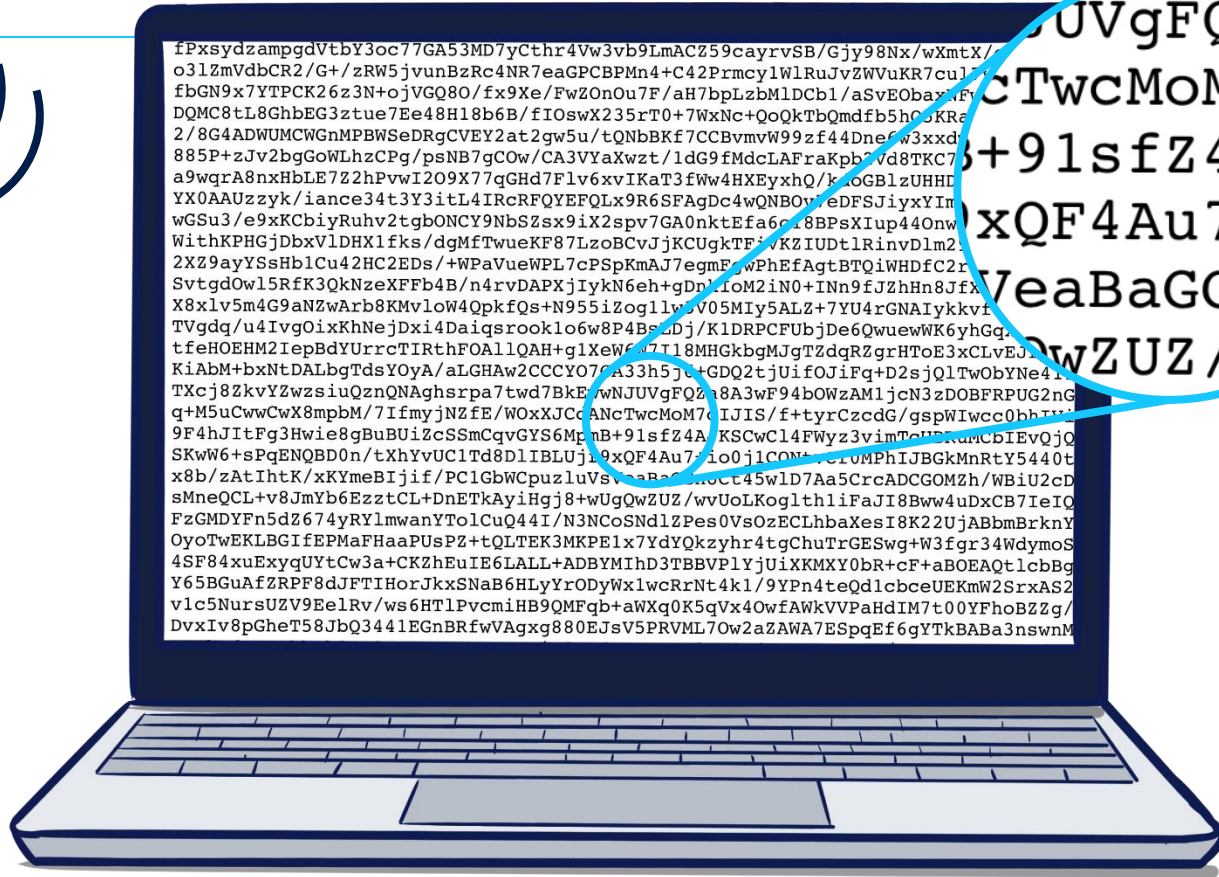
What is Data?



What is data?

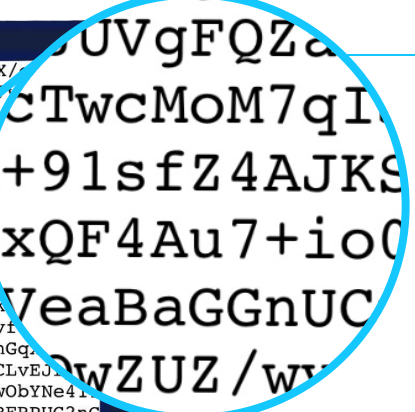
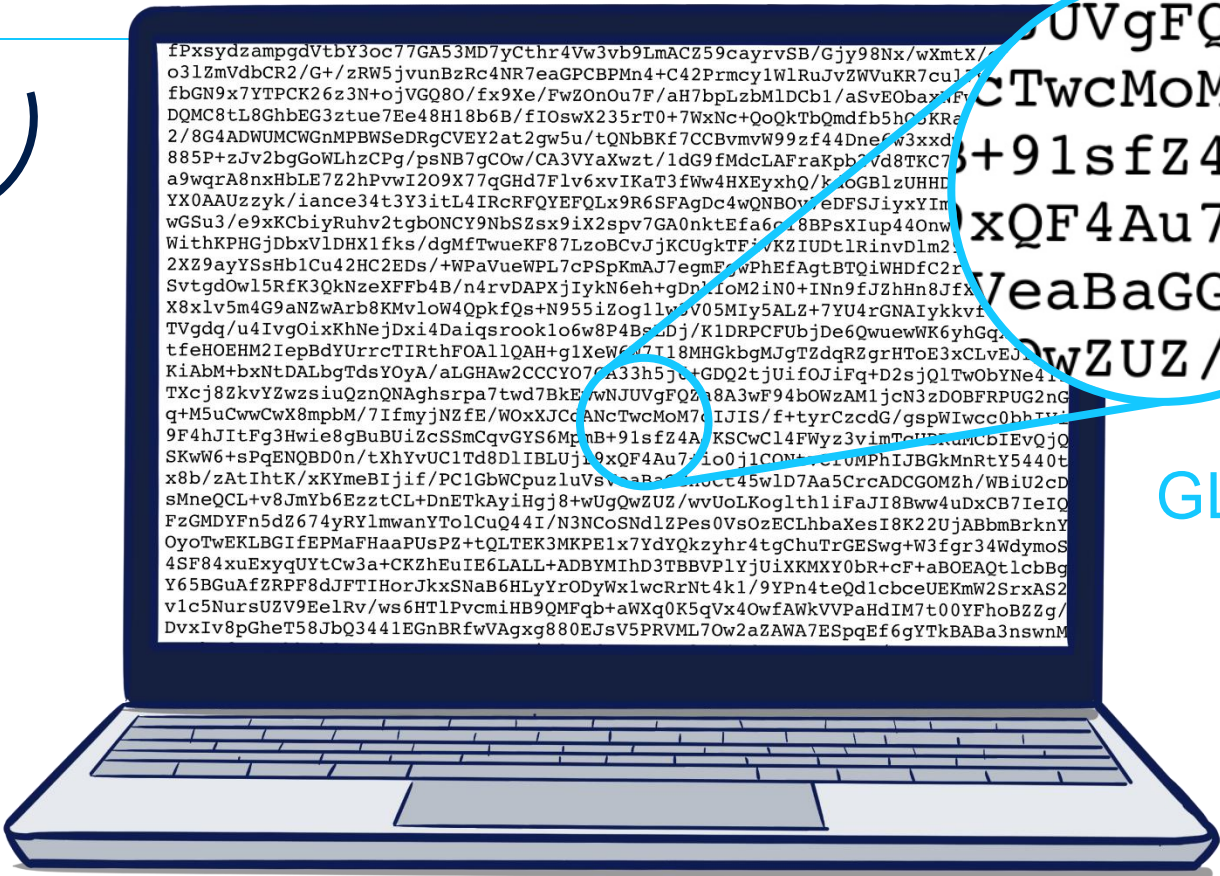


This is data



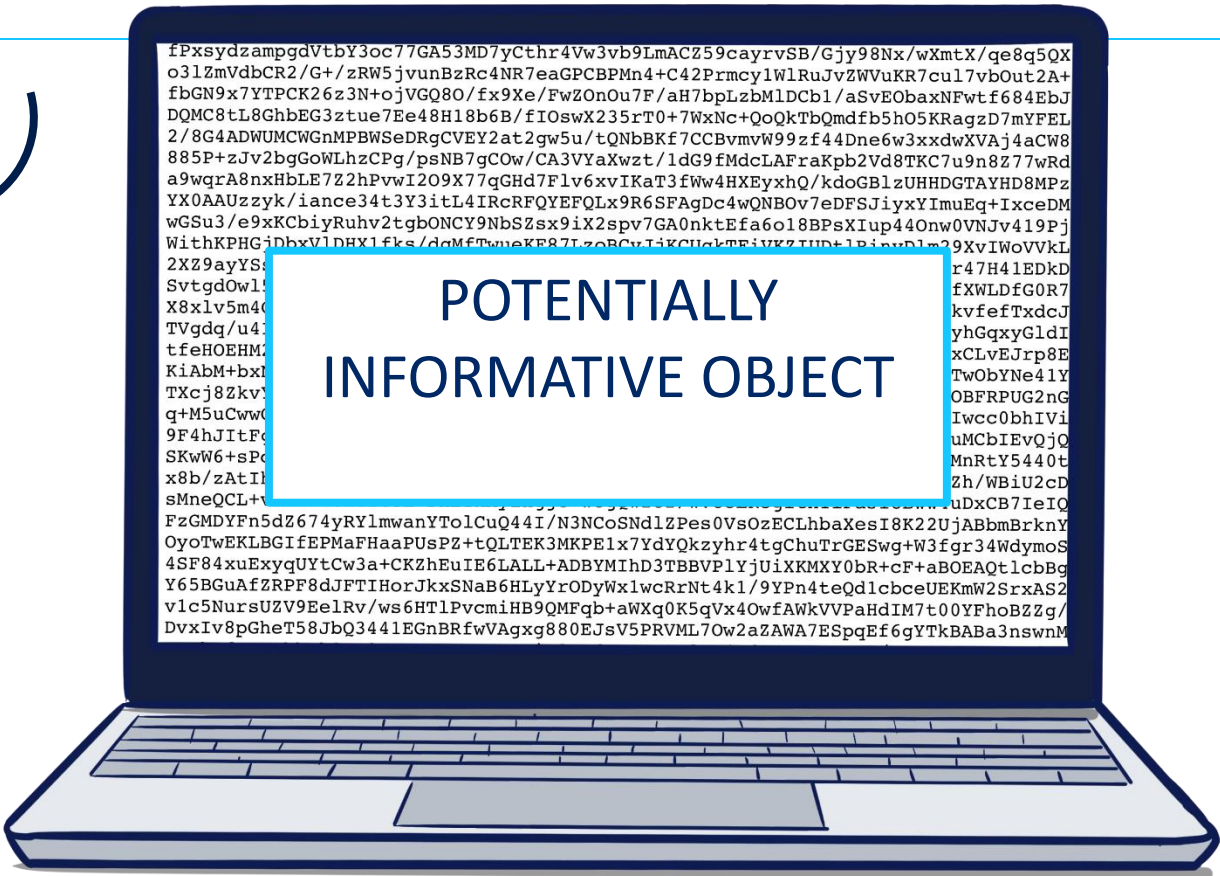
fPxsydzampgdVtbY3oc77GA53MD7yCthr4Vw3vb9LmACZ59cayrvSB/Gjy98Nx/wXmtX/
o31ZmVdbCR2/G+/zRW5jvunBzRc4NR7eaGPCBPMn4+C42Pmcy1WLruJvZWVUKR7cul7y
fbGN9x7YTPCK26z3N+oJVGQ80/fx9Xe/FwZOnOu7F/ah7bpLzBm1DCb1/aSvEObayMfV
DQMC8tL8GhbEG3ztue7Ee48H18b6B/fIOswX235rT0+7WxNc+QoQkTbQmdfb5hcsKRa
2/8G4ADWUMCWGnMPBwSeDRgCVEY2at2gw5u/tQNBbKF7CCBvmvW99zf44DnefW3xxd
885P+zJv2bgGoWLhzCPg/psNB7gCOW/CA3VYAxwzt/ldG9fMdcLAFraKpb7vd8TKC7
a9wqrA8nxHbLE7Z2hPvwI209X77qGhd7Flv6xvIKaT3fww4HXEyxhQ/ktoGB1zUHHD
YX0AAUzzyk/iance34t3Y3itL4IRcRFQYEFQLx9R6SFAgDc4wQNBOv7eDFSJiyxYIn
wGSu3/e9xKChiyRuhv2tgbONCY9NbSzsx9iX2spv7GA0nktEfa6c78BPsXIup440nw
WithKPHGjDbxv1DhX1fks/dgMfTwueKF87LzoBCvJjKCUgkTFvKzIUdt1RinvDlm2
2XZ9ayYSsHb1Cu42HC2EDs/+WPaVueWPL7cPSPkMAJ7egmFwPhEfAgTBTQiWHDfC2r
SvtgdOw15RfK3QkNzeXFFb4B/n4rvDAPXjIyKN6eh+gDn1om2in0+INN9fJzhHn8Jfx
X8xlV5m4G9aNZwArb8KMv1ow4QpkfQs+N955izog1lw5V05MIy5ALz+7YU4rGNAIykkvf
TVgdq/u4IvgOixKhNejdXi4Daiqsrook1o6w8P4Bs2Dj/K1DRPCFUbJDe6QwuewWK6yhGq
tfeHOEHM2IepBdYUrrcTIRthFOAllQAH+g1XeW6w7118MHGkbgMJgTzdqzgrHToE3xCLvEJ
KiAbM+bxNtDALbgTdsYOya/aLGHaw2CCCYO7cA33h5jG+GDQ2tjUiFoJiFq+D2sjQ1TWobYNe41
TXcj8ZkvYZwzsiuQznQNAghsrpa7twd7BkEwNJVUGFQZa8A3wF94bOWzAM1jcN3zDOBFPUg2Ng
q+M5uCwwCwX8mpbM/7IfmyjnzfE/WOxXJCcANcTwcMoM7aIJIS/f+tyrCzcdg/gspWIwcc0bbTW
9F4hJItFg3Hwie8gBuBUiZcSSmCqvGYS6MpnB+91sfZ4AKSCwC14FWyz3vimTcUkAmCBIEvQjO
SKwW6+spqENQBD0n/tXhYvUC1Td8D1IBLUj19xQF4Au7io0j1CONt1rUMPHJJBgkMnRtY5440t
x8b/zAtIhtK/xKYmeBIjif/PC1GbWCpuzluVsVeaB85mCct45w1d7Aa5CrcADCGOMzh/WBiU2cD
smneQCL+v8JmYb6EzztCL+DnETkAyiHgJ8+wUgQwZUZ/wvUoLKog1th1iFaJI8Bww4uDxCB7IEIO
FzGMDYFn5dZ674yRYlmwanYToLCuQ44I/N3NCoSndLzPes0VsOzECLhbaXesI8K22UjABbmBrknY
OyoTweKLBGIfEPMaFHaapUSPz+tQLTEK3MKPE1x7YdYQkzyhr4tgChuTrGESwg+W3fgr34WdymoS
4SF84xuExyQUYtCw3a+CKZhEuIE6LALL+ADBYMThD3TBBVPLyJUiXKMXYObr+cf+aBOEAQt1cbBg
Y65BGuAfZRRPF8dJFTIHOrJkxSNaB6HLyYrODyWx1wCRRnt4k1/9YPn4teQd1cbceUEKmw2SrxAS2
V1c5NursUZV9Ee1Rv/ws6HT1PvcmiHB9QMFqb+aWxQ0K5qVx4OwfAWkVVPaHdIM7t00YFhoBZZg/
DvxIv8pGheT58JbQ3441EGnBRfWVAgxg880EJSv5PRVML7Ow2aZAWA7ESpQef6gyYtkBABA3nswM

This is data



GLYPHS

This is data



POTENTIALLY
INFORMATIVE OBJECT

fPxsydzampgdVtbY3oc77GA53MD7yCthr4Vw3vb9LmACZ59cayrvSB/Gjy98Nx/wXmtX/qe8q5QX
o31ZmVdbCR2/+G/+zRW5jvunBzRc4NR7eaGPCBPMn4+C42PrmcylW1RuJvZwVuKR7cul7vbOut2A+
fbGN9x7YTPCK26z3N+oJVGQ80/fx9Xe/FwZOnOu7F/ah7bpLzbMlDCh1/aSvEObaxNFwt6f684EbJ
DQMC8tL8GhbEG3ztue7Ee48H18b6B/fIOswX235rT0+7WxNc+QoQkTbQmdfb5h05KRagzD7mYFEL
2/8G4ADWUMCWGnMPBWSeDRgCVEY2at2gw5u/tQNbBKf7CCBvmvW99zf44Dne6w3xxdwXVAj4aCW8
885P+zJv2bgGoWlhZCPg/psNB7gCOW/CA3VYaXwzt/lDg9fMdcLAFraKpb2Vd8TKC7u9n8Z77wRd
a9wqrA8nxHbLE7Z2hPvwI209X77qGHd7F1v6xvIKaT3fww4HXEyXhQ/kdoGB1zUHHDGTAYHD8MPz
YX0AAUzzyk/iance34t3Y3itL4IRcRFQYEPQLx9R6SFAgDc4wQNBov7eDFSJiyxYImuEq+IxcDM
wGSu3/e9xKChiyRuhv2tgbONCY9NbSzsx9iX2spv7GA0nktEfa6o18BPsXIup44Onw0VNJv419Pj
WithKPHGjDhxV1DhX1fks/dcmFtweKfE87Lz0RQvLkKChakTEiVKZTUDt1BiyvDlm79XvIWoVvKl
2XZ9ayYSr47H41EDkd
SvtgdOwlfxWLDfGQR7
X8xlv5m40kvfefTxdcJ
TVgdq/u4yhGqxyGLdI
tfeHOEHMxCLvEJrp8E
KiAbM+bxTwObYNe41Y
TXcj8ZkvOBFRPUG2Ng
q+M5uCwW0Iwcc0bhIVI
9F4hJitFumCbIEvQjQ
SKwW6+sPcMnRtY5440t
x8b/zAtIhZh/WBiU2cd
sMneQCL+vuuDxCB7IeIO

FzGMDYFn5dZ674yRYlmwanYTolCuQ44I/N3NCoSnd1ZPes0VsOzECLhbaXesI8K22UjABbmBrknY
OyoTweKLBGIfEPMaFHaapUsPz+tlTEK3MKPE1x7YdyQkzyhr4tgChuTrGESwg+W3fgr34WdymoS
4SF84xuExyqUYtCw3a+CKZhEuIE6LALL+ADBYMThd3TBBVPLyjiXKMXy0br+cF+aBOEAQt1cbBg
Y65BGuaFzRPF8dJFTIHorJkxSNaB6HLyYrODyWxlwCRRnt4k1/9YPn4teQd1cbceUEKmw2SrxAS2
v1c5NursUZV9EelRv/ws6HT1PvcmiHB9QMFqb+aWxcQ0K5qVx4OwfAwkVVPaHdIM7t00YFhoBZzG/
DvxIv8pGheT58JbQ3441EGnBRfVAgxg880EjsV5PRVML7Ow2aZAWA7ESpqEf6gyTkBABA3nswm

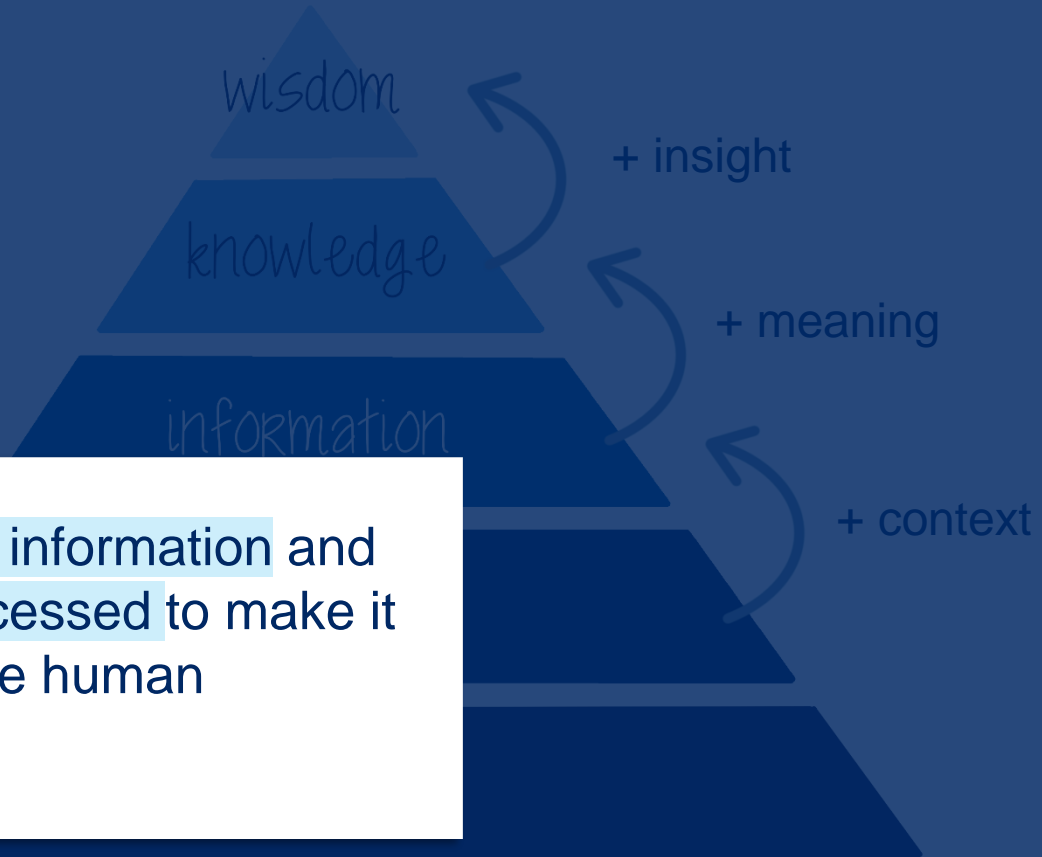
Information – the human-readable data



Knowledge – Where is the cathedral?



Information pyramid



Data is potential information and needs to be processed to make it accessible for the human audience.

Example data - What it really shows

Time

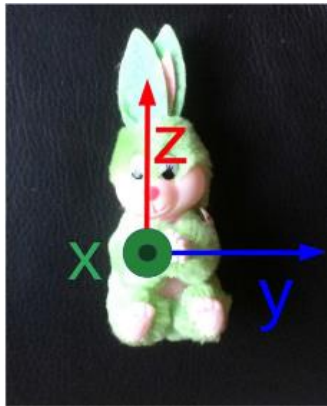


	A	B	C	D	E
1	t	ax	ay	az	scr
2	0	0.3931848	-0.1593144	-0.4178079	0
3	0.01	0.3957354	-0.15696	-0.4242825	0
4	0.04	0.4138839	-0.1547037	-0.429678	0
5	0.05	0.4415481	-0.1512702	-0.4325229	0
6	0.06	0.4741173	-0.1488177	-0.434583	0
7	0.08	0.5021739	-0.1521531	-0.4285008	0
8	0.1	0.5247369	-0.1669662	-0.420849	0
9	0.11	0.5421987	-0.1813869	-0.4160421	0
10	0.14	0.5506353	-0.1947285	-0.4094694	0
11	0.15	0.5538726	-0.203067	-0.4057416	0
12	0.16	0.5534802	-0.2035575	-0.4056435	0
13	0.17	0.5527935	-0.1961019	-0.4098618	0
14	0.2	0.558189	-0.1908045	-0.4121181	0
15	0.21	0.5764356	-0.1865862	-0.4162383	0
16	0.22	0.589581	-0.18639	-0.4258521	0
17	0.25	0.6049827	-0.1941399	-0.4243806	0
18	0.26	0.619992	-0.206991	-0.4192794	0
19	0.27	0.6320583	-0.2191554	-0.4092732	0
20	0.3	0.6392196	-0.2279844	-0.3975993	0
21	0.31	0.6465771	-0.2317122	-0.3908304	0
22	0.32	0.6583491	-0.2291616	-0.3950487	0
23	0.34	0.6725736	-0.2220984	-0.4050549	0

Example data - What it really shows

Time

Biomechanical
acceleration



	A	B	C	D	E
1	t	ax	ay	az	scr
2	0	0.3931848	-0.1593144	-0.4178079	0
3	0.01	0.3957354	-0.15696	-0.4242825	0
4	0.04	0.4138839	-0.1547037	-0.429678	0
5	0.05	0.4415481	-0.1512702	-0.4325229	0
6	0.06	0.4741173	-0.1488177	-0.434583	0
7	0.08	0.5021739	-0.1521531	-0.4285008	0
8	0.1	0.5247369	-0.1669662	-0.420849	0
9	0.11	0.5421987	-0.1813869	-0.4160421	0
10	0.14	0.5506353	-0.1947285	-0.4094694	0
11	0.15	0.5538726	-0.203067	-0.4057416	0
12	0.16	0.5534802	-0.2035575	-0.4056435	0
13	0.17	0.5527935	-0.1961019	-0.4098618	0
14	0.2	0.558189	-0.1908045	-0.4121181	0
15	0.21	0.5764356	-0.1865862	-0.4162383	0
16	0.22	0.589581	-0.18639	-0.4258521	0
17	0.25	0.6049827	-0.1941399	-0.4243806	0
18	0.26	0.619992	-0.206991	-0.4192794	0
19	0.27	0.6320583	-0.2191554	-0.4092732	0
20	0.3	0.6392196	-0.2279844	-0.3975993	0
21	0.31	0.6465771	-0.2317122	-0.3908304	0
22	0.32	0.6583491	-0.2291616	-0.3950487	0
23	0.34	0.6725736	-0.2220984	-0.4050549	0

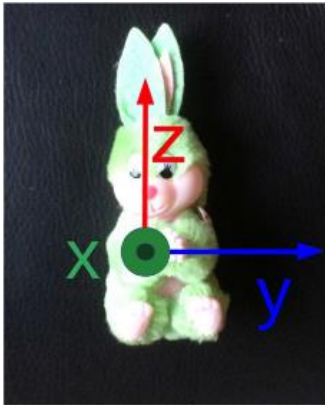
Pendrill, A.-M., Eager, D.(2020). "Velocity, acceleration, jerk, snap and vibration: forces in our bodies during a roller coaster ride." *Phys. Educ.* 55 065012

<https://apparentlysew.weebly.com/stem-blog/category/roller-coaster>

Example data - What it really shows

Time

Biomechanical
acceleration



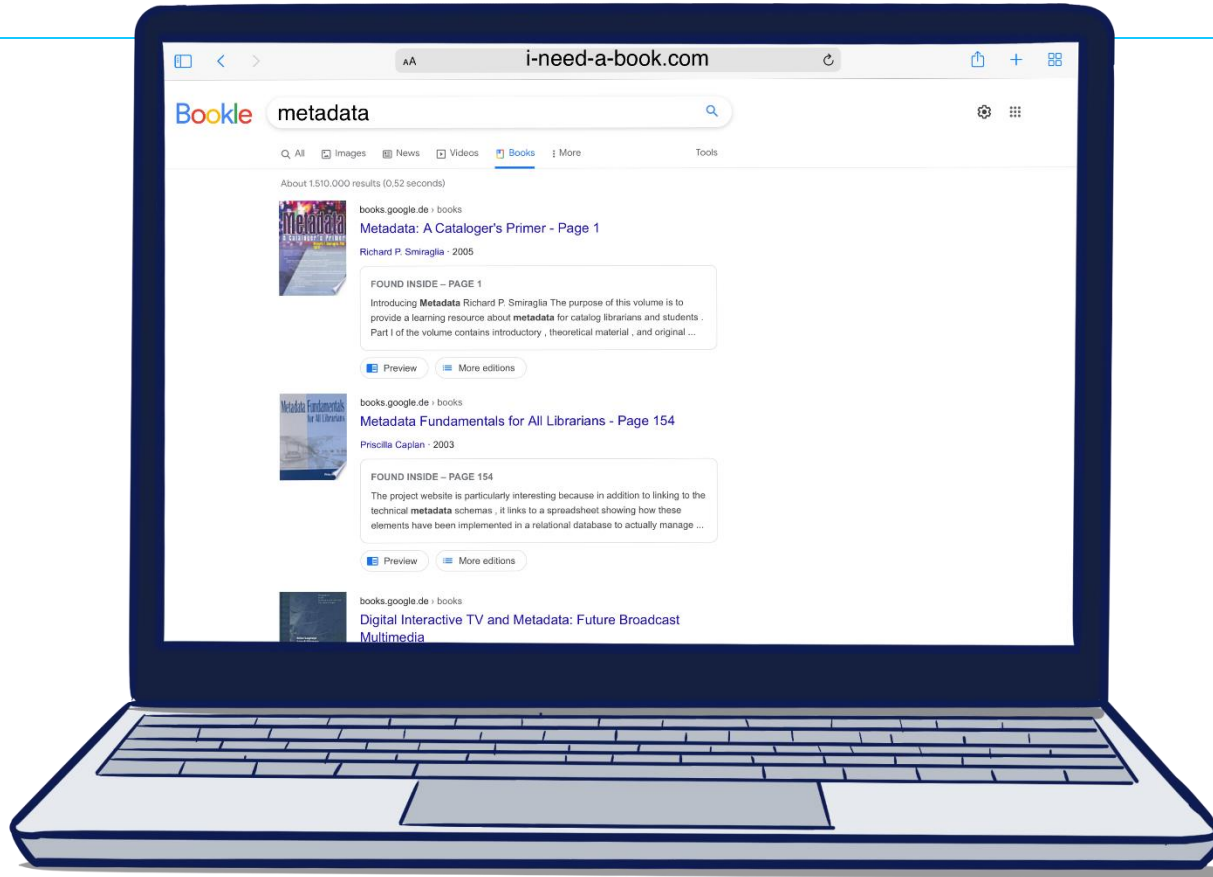
	A	B	C	D	E
1	t	ax	ay	az	scr
2	0	0.3931848	-0.1593144	-0.4178079	0
3	0.01	0.3957354	-0.15696	-0.4242825	0
4	0.04	0.4138839	-0.1547037	-0.429678	0
5	0.05	0.4415481	-0.1512702	-0.4325229	0
6	0.06	0.4741173	-0.1488177	-0.434583	0
7	0.08	0.5021739	-0.1521531	-0.4285008	0
8	0.1	0.5247369	-0.1669662	-0.420849	0
9	0.11	0.5421987	-0.1813869	-0.4160421	0
10	0.14	0.5506353	-0.1947285	-0.4094694	0
11	0.15	0.5538726	-0.203067	-0.4057416	0
12	0.16	0.5534802	-0.2035575	-0.4056435	0
13	0.17	0.5527935	-0.1961019	-0.4098618	0
14	0.2	0.558189	-0.1908045	-0.4121181	0
15	0.21	0.5764356	-0.1865862	-0.4162383	0
16	0.22	0.589581	-0.18639	-0.4258521	0
17	0.25	0.6049827	-0.1941399	-0.4243806	0
18	0.26	0.619992	-0.206991	-0.4192794	0
19	0.27	0.6320583	-0.2191554	-0.4092732	0
20	0.3	0.6392196	-0.2279844	-0.3975993	0
21	0.31	0.6465771	-0.2317122	-0.3908304	0
22	0.32	0.6583491	-0.2291616	-0.3950487	0
23	0.34	0.6725736	-0.2220984	-0.4050549	0

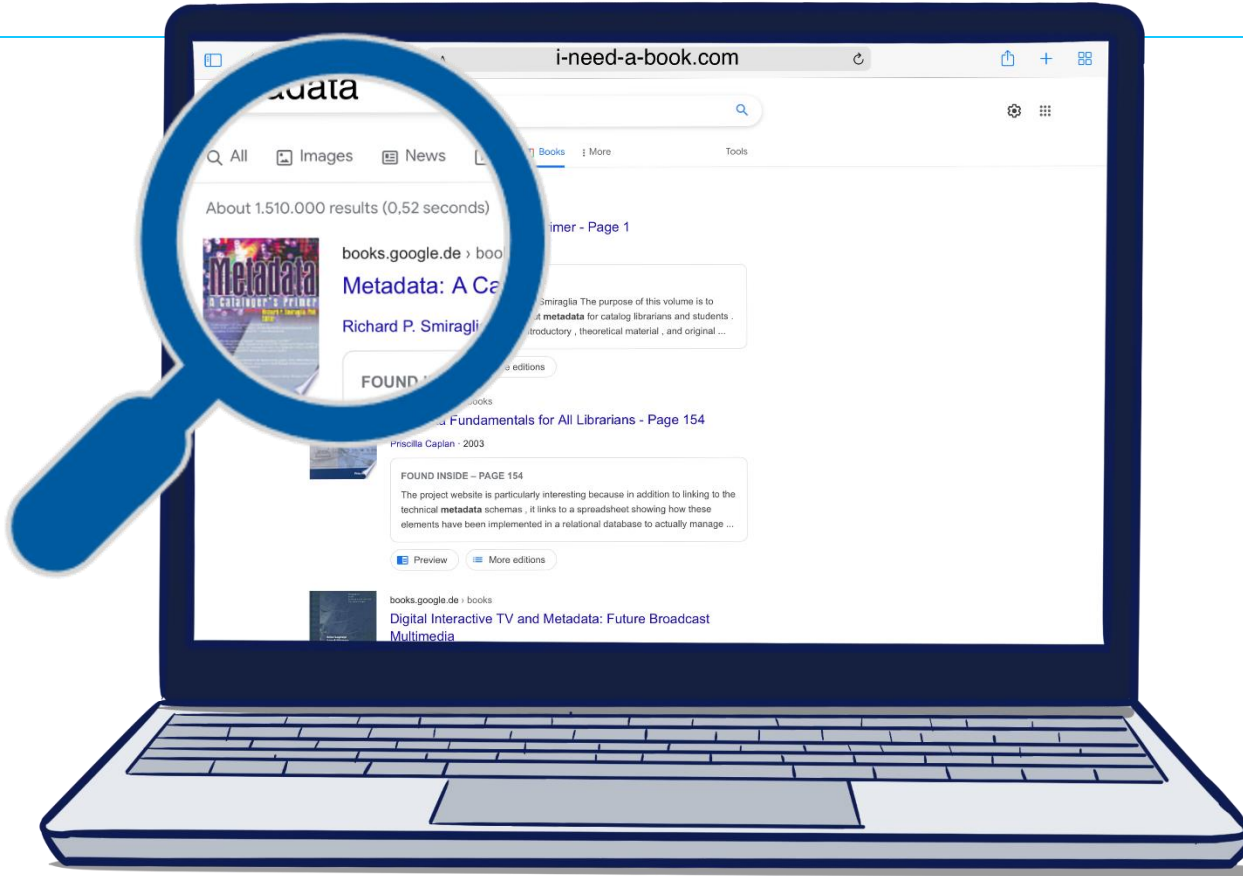
Scream
detected

Pendrill, A.-M., Eager, D.(2020). "Velocity, acceleration, jerk, snap and vibration: forces in our bodies during a roller coaster ride." *Phys. Educ.* 55 065012

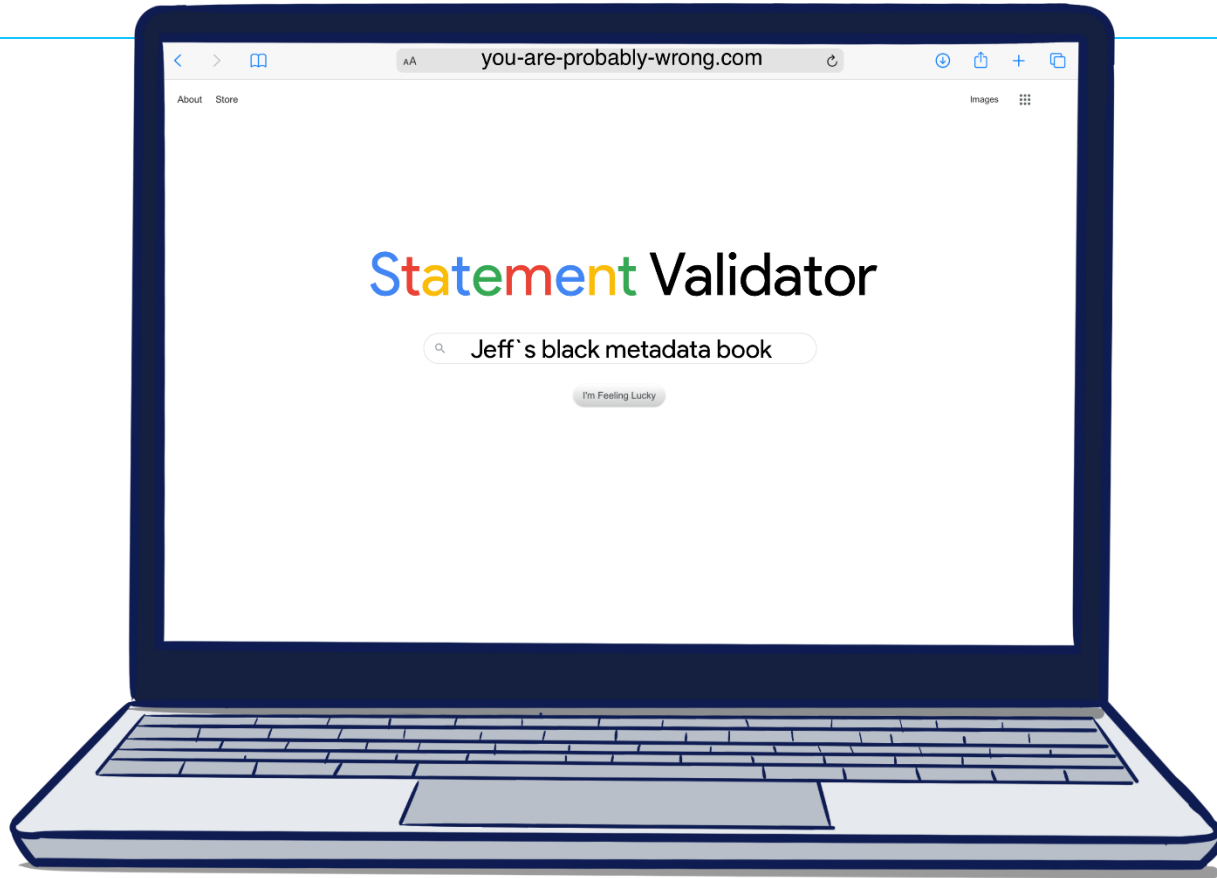
What is Metadata?

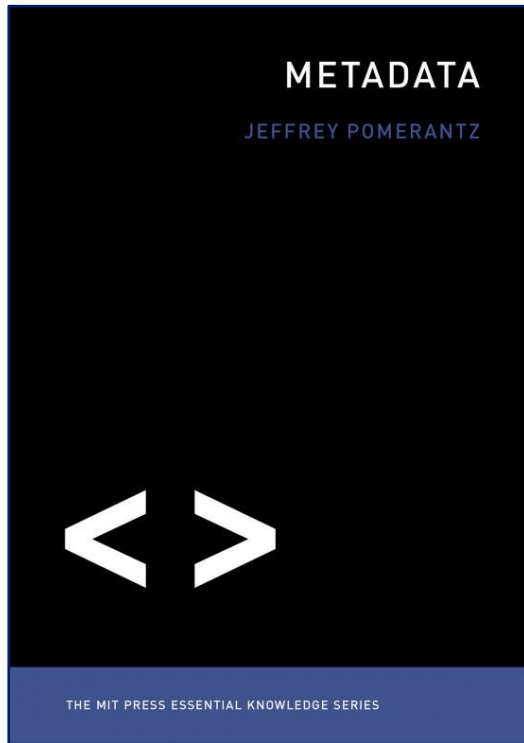




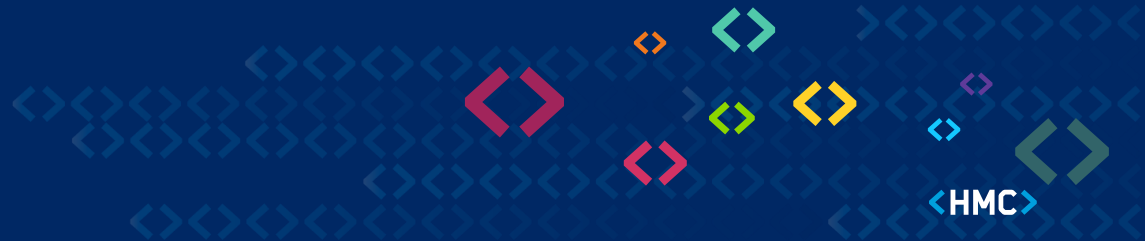








Descriptive, administrative, structural metadata



CONTENTS

© 2015 Jeffrey Pomerantz

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the publisher.

MIT Press books may be purchased at special quantity discounts for business or sales promotional use. For information, please email special_sales@mitpress.mit.edu or write to Special Sales Department, The MIT Press, 1 Rogers Street, Cambridge, MA 02142.

Set in Chaparral Pro by the MIT Press. Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data
Pomerantz, Jeffrey, author.

Metadata / Jeffrey Pomerantz.

pages cm.—(The MIT Press essential knowledge series)

Includes bibliographical references and index.

ISBN 978-0-262-52851-1 (pbk. : alk. paper) 1. Metadata. 2. Information organisation. I. Title.
Z666.7.P66 2015
025.3—dc23

2015030578

10 9 8 7 6 5

Preface ix

- 1 Introduction 1
- 2 Definitions 19
- 3 Descriptive Metadata 65
- 4 Administrative Metadata 93
- 5 Use Metadata 117
- 6 Enabling Technologies for Metadata 133
- 7 The Semantic Web 153
- 8 The Future of Metadata 187

Glossary 209

- Further Readings 215
Figure Credits 223
Bibliography 225
Index 233

Description

- Publication year
- Author
- Title
- Publisher / Series
- Keywords
- Persistent Identifier

© 2015 Jeffrey Pomerantz

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the publisher.

MIT Press books may be purchased at special quantity discounts for business or sales promotional use. For information, please email special_sales@mitpress.mit.edu or write to Special Sales Department, The MIT Press, 1 Rogers Street, Cambridge, MA 02142.

Set in Chaparral Pro by the MIT Press. Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data
Pomerantz, Jeffrey, author.

Metadata / Jeffrey Pomerantz.

pages cm.—(The MIT Press essential knowledge series)

Includes bibliographical references and index.

ISBN 978-0-262-52851-1 (pbk. : alk. paper) 1. Metadata. 2. Information organisation. I. Title.

Z666.7.P66 2015

025.3—dc23

10 9 8 7 6 5

2015030578

CONTENTS

Preface ix

- 1 Introduction 1
- 2 Definitions 19
- 3 Descriptive Metadata 65
- 4 Administrative Metadata 93
- 5 Use Metadata 117
- 6 Enabling Technologies for Metadata 133
- 7 The Semantic Web 153
- 8 The Future of Metadata 187

Glossary 209

Further Readings 215

Figure Credits 223

Bibliography 225

Index 233

Description

- Publication year
- Author
- Title
- Publisher / Series
- Keywords
- Persistent Identifier

© 2015 Jeffrey Pomerantz

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the publisher.

MIT Press books may be purchased at special quantity discounts for business or sales promotional use. For information, please email special_sales@mitpress.mit.edu or write to Special Sales Department, The MIT Press, 1 Rogers Street, Cambridge, MA 02142.

Set in Chaparral Pro by the MIT Press. Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data
Pomerantz, Jeffrey, author.

Metadata / Jeffrey Pomerantz.

pages cm.—(The MIT Press essential knowledge series)

Includes bibliographical references and index.

ISBN 978-0-262-52851-1 (pbk. : alk. paper) 1. Metadata. 2. Information

Organization. I. Title.

Z666.7.P66 2015

025.3—dc23

10 9 8 7 6 5

2015030578

CONTENTS

Preface ix

- 1 Introduction 1
- 2 Definitions 19
- 3 Descriptive Metadata 65
- 4 Administrative Metadata 93
- 5 Use Metadata 117
- 6 Enabling Technologies for Metadata 133
- 7 The Semantic Web 153
- 8 The Future of Metadata 187

Glossary 209

Further Readings 215

Figure Credits 223

Bibliography 225

Index 233

Description

- Publication year
- Author
- Title
- Publisher / Series
- Keywords
- Persistent Identifier

© 2015 Jeffrey Pomerantz

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the publisher.

MIT Press books may be purchased at special quantity discounts for business or sales promotional use. For information, please email special_sales@mitpress.mit.edu or write to Special Sales Department, The MIT Press, 1 Rogers Street, Cambridge, MA 02142.

Set in Chaparral Pro by the MIT Press. Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data
Pomerantz, Jeffrey, author.

Metadata / Jeffrey Pomerantz.

pages cm.—(The MIT Press essential knowledge series)

Includes bibliographical references and index.

ISBN 978-0-262-52851-1 (pbk. : alk. paper) 1. Metadata. 2. Information organisation. I. Title.

Z666.7.P66 2015

025.3—dc23

10 9 8 7 6 5

2015030578

CONTENTS

Preface ix

- 1 Introduction 1
- 2 Definitions 19
- 3 Descriptive Metadata 65
- 4 Administrative Metadata 93
- 5 Use Metadata 117
- 6 Enabling Technologies for Metadata 133
- 7 The Semantic Web 153
- 8 The Future of Metadata 187

Glossary 209

Further Readings 215

Figure Credits 223

Bibliography 225

Index 233

Administration

- Distribution
- Responsibility
- Conditions

Description

- Publication year
- Author
- Title
- Publisher / Series
- Keywords
- Persistent Identifier

© 2015 Jeffrey Pomerantz

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the publisher.

MIT Press books may be purchased at special quantity discounts for business or sales promotional use. For information, please email special_sales@mitpress.mit.edu or write to Special Sales Department, The MIT Press, 1 Rogers Street, Cambridge, MA 02142.

Set in Chaparral Pro by the MIT Press. Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data
Pomerantz, Jeffrey, author.

Metadata / Jeffrey Pomerantz.

pages cm.—(The MIT Press essential knowledge series)

Includes bibliographical references and index.

ISBN 978-0-262-52851-1 (pbk. : alk. paper) 1. Metadata. 2. Information organisation. I. Title.

Z666.7.P66 2015

025.3—dc23

10 9 8 7 6 5

2015030578

CONTENTS

Preface ix

- 1 Introduction 1
- 2 Definitions 19
- 3 Descriptive Metadata 65
- 4 Administrative Metadata 93
- 5 Use Metadata 117
- 6 Enabling Technologies for Metadata 133
- 7 The Semantic Web 153
- 8 The Future of Metadata 187

Glossary 209

Further Readings 215

Figure Credits 223

Bibliography 225

Index 233

Administration

- Distribution
- Responsibility
- Conditions

Legal terms

- Copyright issues
- Terms of distribution

Description

- Publication year
- Author
- Title
- Publisher / Series
- Keywords
- Persistent Identifier

© 2015 Jeffrey Pomerantz

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the publisher.

MIT Press books may be purchased at special quantity discounts for business or sales promotional use. For information, please email special_sales@mitpress.mit.edu or write to Special Sales Department, The MIT Press, 1 Rogers Street, Cambridge, MA 02142.

Set in Chaparral Pro by the MIT Press. Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data
Pomerantz, Jeffrey, author.

Metadata / Jeffrey Pomerantz.
pages cm.—(The MIT Press essential knowledge series)
Includes bibliographical references and index.
ISBN 978-0-262-52851-1 (pbk. : alk. paper) 1. Metadata. 2. Information organization. I. Title.
Z666.7.P66 2015
025.3—dc23

10 9 8 7 6 5

2015030578

CONTENTS

Preface ix

1	Introduction	1
2	Definitions	19
3	Descriptive Metadata	65
4	Administrative Metadata	93
5	Use Metadata	117
6	Enabling Technologies for Metadata	133
7	The Semantic Web	153
8	The Future of Metadata	187

Glossary	209
Further Readings	215
Figure Credits	223
Bibliography	225
Index	233

Administration

- Distribution
- Responsibility
- Conditions

Legal terms

- Copyright issues
- Terms of distribution

Structure

- Content
- Chapters
- Pages

Descriptive Metadata:

Metadata that provides information to **identify and describe a resource**, such as its title, author, and keywords.

Administrative Metadata:

Metadata used to **manage a resource**, including details like rights, licensing, and file creation dates.

Structural Metadata:

Metadata that outlines how **a resource is organized**, such as its chapters, sections, or relationships between components.

Description

- Publication year
- Author
- Title
- Publisher / Series
- Keywords
- Persistent Identifier

© 2015 Jeffrey Pomerantz

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the publisher.

MIT Press books may be purchased at special quantity discounts for business or sales promotional use. For information, please email special_sales@mitpress.mit.edu or write to Special Sales Department, The MIT Press, 1 Rogers Street, Cambridge, MA 02142.

CONTENTS

Preface ix

- 1 Introduction 1
- 2 Definitions 19
- 3 Descriptive Metadata 65
- 4 Administrative Metadata 93
- 5 Use Metadata 117
- 6 Technologies for Metadata 133
- 7 Semantic Web 153
- 8 Types of Metadata 187
- 9 209
- 10 Readings 215
- 11 Credits 223
- 12 Bibliography 225
- 13 3

Administration

- Distribution
- Responsibility
- Conditions

Legal terms

- Copyright issues
- Terms of distribution

Structure

- Content
- Chapters
- Pages



METADATA

or metainformation

is structured data that contains information about characteristics of other data (objects).

Questions?



Metadata in a paper

Demonstration





HELMHOLTZ
Metadata Collaboration

Metadata Annotation in the Scientific Context



You should start
your project with
repeating your
collaborator's
results



You should start your project with **repeating** your collaborator's results



The Publication

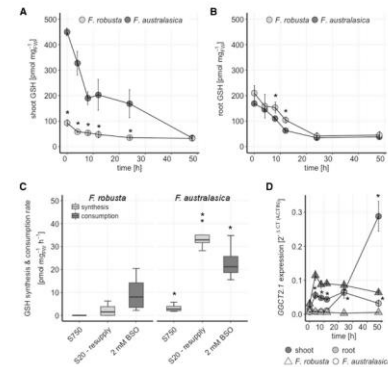


Figure 4. GSH turnover in *F. robusta* and *F. australasica*. GSH concentrations in shoots (A) and roots (B) of 20-d-old seedlings of *F. robusta* (C_1) and *F. australasica* (C_2) in a time course of 48 h after transfer to medium supplemented with 2 mM BSO. Data are presented as means and $n_s = 4$. C, GSH synthesis was analyzed in 20-d-old seedlings exposed to low sulfate (20 μM sulfate; S20) or adequate sulfate (750 μM sulfate; S750) for 4 d by resupply with 0.2 mM ³⁵Sulfate solution for 4 h. GSH consumption rate is calculated from A, at 4 h after treatment with 2 mM BSO. Data are shown as box plot (25%–75%) the line represents mean, and the whiskers represent 1.5 IQR, $n_s = 4$. D, Transcript levels of GSH:TS in shoots and roots of 20-d-old seedlings in a time course of 48 h after transfer to medium supplemented with 2 mM BSO. Data are presented as means and SEM, $n_s = 4$. Asterisks represent significant differences between *F. robusta* and *F. australasica* at $P < 0.05$ (Student's *t* test).

to higher GSH synthesis are therefore likely to be involved in the adjustment of S supply and GSH homeostasis of C_1 plants.

Partitioning of S in Shoots and Roots of *Flaveria* Species

To test the significance of the root for S metabolism in the context of the evolution of C_1 photosynthesis, the five species were grown under full nutrient and low S conditions. Total S, sulfate and low M_s thiols were determined in shoots and roots (Supplemental Fig. S7). Whereas total S and sulfate did not show any clear patterns relative to photosynthetic type, Cys, and GSH

at full nutrition. To better understand the partitioning of S in the different species, the relative portions of total S in sulfate, Cys, and GSH were calculated (Fig. 5). In the shoots of fully nourished *Flaveria* species, the fraction of total S occupied by inorganic sulfate was relatively stable at 50%–70%. However, in the roots, the fraction of inorganic sulfate was higher in the C_1 species. Exposure to S deficiency reduced the sulfate pool in the shoots and roots of *F. robusta*, *F. linearis*, *F. anomala*, and *F. palmeri* to 3.5%–16%. The C_2 species *F. australasica* suffered little loss of relative sulfate pool in shoots, but showed a strong decrease in roots. The increase in GSH fractions of total S in shoots and roots

You should start your project with **repeating** your collaborator's results



The Data

An illustration of a laptop with a large table of numerical data on its screen. The table has 10 columns and 25 rows of data. The data is presented in a grid format with alternating light and dark blue rows.

21.5	21.6	20.8	20.2	20.8	21.0	21.6	20.8	21.2	21.1	
61.3	60.7	44.8	46.2	49.2	49.1	49.3	48.0	40.1	41.3	
18.0	15.8	15.3	14.0	14.4	15.3	15.4	14.6	14.8	14.0	
16.7	16.8	16.3	17.6	18.3	17.6	17.5	18.3	17.9	17.7	
20.2	20.6	20.1	20.0	19.7	19.9	19.6	20.3	20.6	20.0	
22.0	22.0	21.8	23.4	21.7	23.1	23.4	23.5	26.0	24.2	
23.3	23.1	23.7	25.7	27.3	29.4	30.3	29.9	27.5	25.9	
29.3	28.3	28.1	27.6	27.7	31.0	34.6	35.7	36.0	35.7	
24.0	23.3	23.8	24.7	26.1	26.7	27.2	27.3	29.2	28.6	
18.8	19.0	18.5	18.5	19.2	19.3	19.1	18.1	18.5	17.7	
				31.1	32.6	32.6	29.9	29.3	29.1	
25.9	26.0	25.5	24.9	25.0	28.1	29.9	28.5	28.3	28.7	
25.4	25.2	23.3	23.5	24.6	24.6	27.1	27.8	27.4	28.9	
42.2	35.1	34.2	37.9	38.2	40.1	36.2	35.1	32.7	30.9	28.5
35.9	28.7	28.3	29.6	34.0	33.1	32.5	30.8	27.3	29.3	
16.5	15.9	15.5	17.8	17.1	16.8	18.4	19.0	19.0	18.5	
31.4	29.4	28.2	29.6	29.9	31.5	33.5	34.8	31.8	28.2	26.3
19.5	19.7	20.1	20.3	21.2	22.1	23.1	24.0	23.8	22.4	
16.0	15.7	14.9	15.1	15.1	15.7	15.0	15.9	16.5	16.4	
17.8	16.7	20.6	19.1	18.9	19.2	18.5	18.8	19.2	18.3	
39.5	34.4	30.5	27.8	27.8	27.2	26.7	25.8	24.7	23.4	
25.0	25.0	26.0	24.9	25.3	24.4	25.3	27.5	27.5	26.6	
	47.0	44.2	43.0	41.5	40.9	43.2	41.9	40.3	37.4	
17.1	17.1	18.5	17.1	18.3	19.3	19.6	20.4	20.4	19.2	
26.7	21.4	20.6	19.6	20.6	20.6	20.5	19.8	18.4	18.4	
17.1	17.4	17.4	16.9	16.9	17.9	17.2	16.0	17.3	16.8	

Monya Baker

„More than 70 % of researchers have tried and failed to reproduce another scientist’s experiments.

More than half have failed to reproduce their own experiments. “

Quote: Baker, M. 1,500 scientists lift the lid on reproducibility. *Nature* 533, 452 – 454 (2016). <https://doi.org/10.1038/533452a>

Image: <https://www.booksmith.com/event/bindery-launch-katie-burke-urban-playground-what-kids-say-about-living-san-francisco>

Slightly better – lab notes

	A	B	C	D	E
1	t	ax	ay	az	scr
2	0	0.3931848	-0.1593144	-0.4178079	0
3	0.01	0.3957354	-0.15696	-0.4242825	0
4	0.04	0.4138839	-0.1547037	-0.429678	0
5	0.05	0.4415481	-0.1512702	-0.4325229	0
6	0.06	0.4741173	-0.1488177	-0.434583	0
7	0.08	0.5021739	-0.1521531	-0.4285008	0
8	0.1	0.5247369	-0.1669662	-0.420849	0
9	0.11	0.5421987	-0.1813869	-0.4160421	0
10	0.14	0.5506353	-0.1947285	-0.4094694	0
11	0.15	0.5538726	-0.203067	-0.4057416	0
12	0.16	0.5534802	-0.2035575	-0.4056435	0
13	0.17	0.5527935	-0.1961019	-0.4098618	0
14	0.2	0.558189	-0.1908045	-0.4121181	0
15	0.21	0.5764356	-0.1865862	-0.4162383	0
16	0.22	0.589581	-0.18639	-0.4258521	0
17	0.25	0.6049827	-0.1941399	-0.4243806	0
18	0.26	0.619992	-0.206991	-0.4192794	0
19	0.27	0.6320583	-0.2191554	-0.4092732	0
20	0.3	0.6392196	-0.2279844	-0.3975993	0
21	0.31	0.6465771	-0.2317122	-0.3908304	0
22	0.32	0.6583491	-0.2291616	-0.3950487	0
23	0.34	0.6725736	-0.2220984	-0.4050549	0



someRandomFileName.csv

Slightly better – lab notes

	A	B	C	D	E
1	t	ax	ay	az	scr
2	0	0.3931848	-0.1593144	-0.4178079	0
3	0.01	0.3957354	-0.15696	-0.4242825	0
4	0.04	0.4138839	-0.1547037	-0.429678	0
5	0.05	0.4415481	-0.1512702	-0.4325229	0
6	0.06	0.4741173	-0.1488177	-0.434583	0
7	0.08	0.5021739	-0.1521531	-0.4285008	0
8	0.1	0.5247369	-0.1669662	-0.420849	0
9	0.11	0.5421987	-0.1813869	-0.4160421	0
10	0.14	0.5506353	-0.1947285	-0.4094694	0
11	0.15	0.5538726	-0.203067	-0.4057416	0
12	0.16	0.5534802	-0.2035575	-0.4056435	0
13	0.17	0.5527935	-0.1961019	-0.4098618	0
14	0.2	0.558189	-0.1908045	-0.4121181	0
15	0.21	0.5764356	-0.1865862	-0.4162383	0
16	0.22	0.589581	-0.18639	-0.4258521	0
17	0.25	0.6049827	-0.1941399	-0.4243806	0
18	0.26	0.619992	-0.206991	-0.4192794	0
19	0.27	0.6320583	-0.2191554	-0.4092732	0
20	0.3	0.6392196	-0.2279844	-0.3975993	0
21	0.31	0.6465771	-0.2317122	-0.3908304	0
22	0.32	0.6583491	-0.2291616	-0.3950487	0
23	0.34	0.6725736	-0.2220984	-0.4050549	0



20220228_recordingData.csv

2022-02-28

Gotham City, New Jersey, USA

Flight of the bat

weather: more clouds than sun, 11°C, 74% humidity,

1023 mbar, SSW, 17 km/h

recording device strapped to upper arm

Slightly better – lab notes

	A	B	C	D	E
1	t	ax	ay	az	scr
2	0	0.3931848	-0.1593144	-0.4178079	0
3	0.01	0.3957354	-0.15696	-0.4242825	0
4	0.04	0.4138839	-0.1547037	-0.429678	0
5	0.05	0.4415481	-0.1512702	-0.4325229	0
6	0.06	0.4741173	-0.1488177	-0.434583	0
7	0.08	0.5021739	-0.1521531	-0.4285008	0
8	0.1	0.5247369	-0.1669662	-0.420849	0
9	0.11	0.5421987	-0.1813869	-0.4160421	0
10	0.14	0.5506353	-0.1947285	-0.4094694	0
11	0.15	0.5538726	-0.203067	-0.4057416	0
12	0.16	0.5534802	-0.2035575	-0.4056435	0
13	0.17	0.5527935	-0.1961019	-0.4098618	0
14	0.2	0.558189	-0.1908045	-0.4121181	0
15	0.21	0.5764356	-0.1865862	-0.4162383	0
16	0.22	0.589581	-0.18639	-0.4258521	0
17	0.25	0.6049827	-0.1941399	-0.4243806	0
18	0.26	0.619992	-0.206991	-0.4192794	0
19	0.27	0.6320583	-0.2191554	-0.4092732	0
20	0.3	0.6392196	-0.2279844	-0.3975993	0
21	0.31	0.6465771	-0.2317122	-0.3908304	0
22	0.32	0.6583491	-0.2291616	-0.3950487	0
23	0.34	0.6725736	-0.2220984	-0.4050549	0



20220228_recordingData.csv

LAB NOTES IV

LAB NOTES III

LAB NOTES II

LAB NOTES I

2022-02-28

Gotham City, New Jersey, USA

Flight of the bat

weather: more clouds than sun, 11°C, 74% humidity,

1023 mbar, SSW, 17 km/h

recording device strapped to upper arm

Slightly better – lab notes

	A	B		
1	t	ax		
2	0	0.3931		
3	0.01	0.3957		
4	0.04	0.4138		
5	0.05	0.4415481	-0.1512702	-0.4325229
6	0.06	0.4741173	-0.1488177	-0.434583
7	0.08	0.5021		
8	0.1	0.5247		
9	0.11	0.5421		
10	0.14	0.5506		
11	0.15	0.5538726	-0.1	
12	0.16	0.5534802	-0.2	
13	0.17	0.5527935	-0.1	
14	0.2	0.558189	-0.1	
15	0.21	0.5764356	-0.1	
16	0.22	0.589581	-0.1	
17	0.25	0.6049827	-0.1	
18	0.26	0.619992	-0.2	
19	0.27	0.6320583	-0.2191554	-0.4092732
20	0.3	0.6392196	-0.2279844	-0.3975993
21	0.31	0.6465771	-0.2317122	-0.3908304
22	0.32	0.6583491	-0.2291616	-0.3950487
23	0.34	0.6725736	-0.2220984	-0.4050549



- some kind of documentation



- unstructured
- hard to find
- seperated from data
- hard to share / only in the possession of the experimentator
- frequently hard to read

dingData.csv

LAB NOTES IV
LAB NOTES III

2022-02-28

11°C, 76% humidity

recording device strapped to upper arm

Gregor Mendel, the father of genetics

Do you know why Mendel's work was largely **ignored** until 1900, **more than 30 years** after his death?

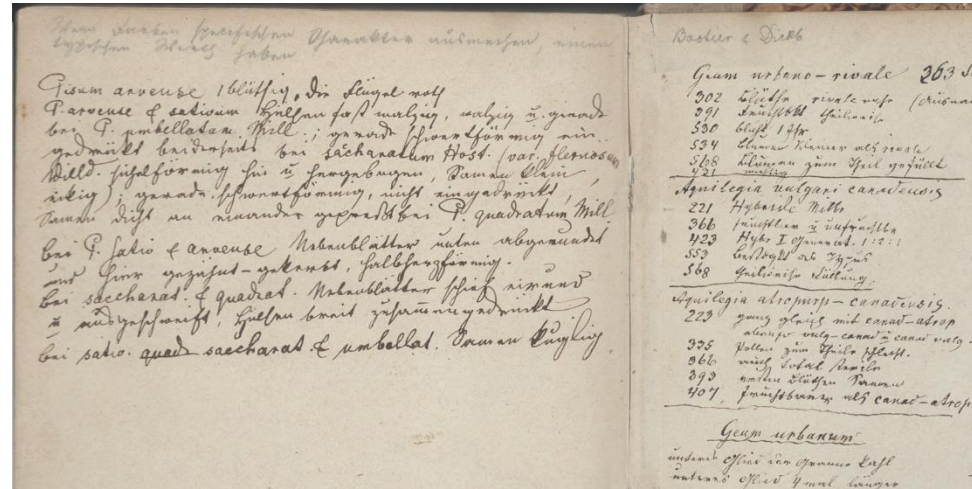


Other scientists **couldn't fully understand and validate** his findings

Mendel did **not** provide sufficient details in his notes about

- his experimental design or
- some of the intermediate steps in his calculations

Without this detailed “**metadata**”, it was difficult for other scientists to **replicate his work** or even **fully appreciate its significance**.



<https://www.nature.com/articles/s41588-022-01109-9/figures/1>

Even better – Readme style metadata

	A	B	C	D	E
1	<u>t</u>	<u>ax</u>	<u>ay</u>	<u>az</u>	<u>scr</u>
2	0	0.3931848	-0.1593144	-0.4178079	0
3	0.01	0.3957354	-0.15696	-0.4242825	0
4	0.04	0.4138839	-0.1547037	-0.429678	0
5	0.05	0.4415481	-0.1512702	-0.4325229	0
6	0.06	0.4741173	-0.1488177	-0.434583	0
7	0.08	0.5021739	-0.1521531	-0.4285008	0
8	0.1	0.5247369	-0.1669662	-0.420849	0
9	0.11	0.5421987	-0.1813869	-0.4160421	0
10	0.14	0.5506353	-0.1947285	-0.4094694	0
11	0.15	0.5538726	-0.203067	-0.4057416	0
12	0.16	0.5534802	-0.2035575	-0.4056435	0
13	0.17	0.5527935	-0.1961019	-0.4098618	0
14	0.2	0.558189	-0.1908045	-0.4121181	0
15	0.21	0.5764356	-0.1865862	-0.4162383	0
16	0.22	0.589581	-0.18639	-0.4258521	0
17	0.25	0.6049827	-0.1941399	-0.4243806	0
18	0.26	0.619992	-0.206991	-0.4192794	0
19	0.27	0.6320583	-0.2191554	-0.4092732	0
20	0.3	0.6392196	-0.2279844	-0.3975993	0
21	0.31	0.6465771	-0.2317122	-0.3908304	0
22	0.32	0.6583491	-0.2291616	-0.3950487	0
23	0.34	0.6725736	-0.2220984	-0.4050549	0



20220228_recordingData.csv

Even better – Readme style metadata

	A	B	C	D	E
1	<u>t</u>	<u>ax</u>	<u>ay</u>	<u>az</u>	<u>scr</u>
2	0	0.3931848	-0.1593144	-0.4178079	0
3	0.01	0.3957354	-0.15696	-0.4242825	0
4	0.04	0.4138839	-0.1547037	-0.429678	0
5	0.05	0.4415481	-0.1512702	-0.4325229	0
6	0.06	0.4741173	-0.1488177	-0.434583	0
7	0.08	0.5021739	-0.1521531	-0.4285008	0
8	0.1	0.5247369	-0.1669662	-0.420849	0
9	0.11	0.5421987	-0.1813869	-0.4160421	0
10	0.14	0.5506353	-0.1947285	-0.4094694	0
11	0.15	0.5538726	-0.203067	-0.4057416	0
12	0.16	0.5534802	-0.2035575	-0.4056435	0
13	0.17	0.5527935	-0.1961019	-0.4098618	0
14	0.2	0.558189	-0.1908045	-0.4121181	0
15	0.21	0.5764356	-0.1865862	-0.4162383	0
16	0.22	0.589581	-0.18639	-0.4258521	0
17	0.25	0.6049827	-0.1941399	-0.4243806	0
18	0.26	0.619992	-0.206991	-0.4192794	0
19	0.27	0.6320583	-0.2191554	-0.4092732	0
20	0.3	0.6392196	-0.2279844	-0.3975993	0
21	0.31	0.6465771	-0.2317122	-0.3908304	0
22	0.32	0.6583491	-0.2291616	-0.3950487	0
23	0.34	0.6725736	-0.2220984	-0.4050549	0



20220228_recordingData.csv



20220228_recordingData_Readme.txt

```
Open [v] [f] 20220228_trainingObject_Readme.txt -/Documents/AS-2/HMC/HubInfo_Tr...ofScientificMetadata/material Save [≡] - [x]
1 trainingobject.csv
2
3
4 The data describes the biomechanical acceleration and screams detected of a test person during
5 the ride of the roller coaster "Flight of the Bat" in Gotham City.
6
7 The data was collected by Bruce Wayne and Selina Kyle (Institute for Vigilance and Nightly Motion
8 - Justice League) on 2022-02-28 in Gotham City, New Jersey.
9 Weather conditions were optimal for the measurement, 11°C, more clouds than sun, 74% humidity,
10 SSM wind with 17 km/h velocity.
11
12 Test person:
13 The test person (male) is 5'11 tall and weighs 187 lbs.
14
15 Recording procedure:
16 The test person strapped the recording device (iPhone X) with a running armband to the left upper
17 arm and activated the biomechanical acceleration and screen detection of the application Physics
18 Toolbox Suite by Vleyra Software.
19 During the ride, the test person was instructed to grasp the seat handles tightly to avoid
20 excessive movement of the arm. The test person was seated in row 10 on the outer left (seat 37).
21
22 Recorded variables:
23 "t" describes the ride time at which measurements were taken upon activating the recording.
24 "ax" describes the biomechanical acceleration of the test person on the x axis in m/s2.
25 "ay" describes the biomechanical acceleration of the test person on the y axis in m/s2.
26 "az" describes the biomechanical acceleration of the test person on the z axis in m/s2.
27 "scr" is a boolean indicating a detected scream of the test person.]
```

Even better – Readme style metadata

	A	B	C	D	E
1	t	ax	ay	az	scr
2	0	0.3931848	-0.1593144	-0.4178079	0
3	0.01	0.3957354	-0.15696	-0.4242825	0
4	0.04	0.4138839	-0.1547037	-0.429678	0
5	0.05	0.4415481	-0.1512702	-0.4325229	0
6	0.06	0.4741173	-0.1488177	-0.434583	0
7	0.08	0.5021739	-0.1521531	-0.4285008	0
8	0.1	0.5247369	-0.1669662	-0.420849	0
9	0.11	0.5421987	-0.1813869	-0.4160421	0
10	0.14	0.5506353	-0.1947285	-0.4094694	0
11	0.15	0.5538726	-0.203067	-0.4057416	0
12	0.16	0.5534802	-0.2035575	-0.4056435	0
13	0.17	0.5527935	-0.1961019	-0.4098618	0
14	0.2	0.558189	-0.1908045	-0.4121181	0
15	0.21	0.5764356	-0.1865862	-0.4162383	0
16	0.22	0.589581	-0.18639	-0.4258521	0
17	0.25	0.6049827	-0.1941399	-0.4243806	0
18	0.26	0.619992	-0.206991	-0.4192794	0
19	0.27	0.6320583	-0.2191554	-0.4092732	0
20	0.3	0.6392196	-0.2279844	-0.3975993	0
21	0.31	0.6465771	-0.2317122	-0.3908304	0
22	0.32	0.6583491	-0.2291616	-0.3950487	0
23	0.34	0.6725736	-0.2220984	-0.4050549	0



20220228_recordingData.csv



20220228_recordingData_Readme.txt

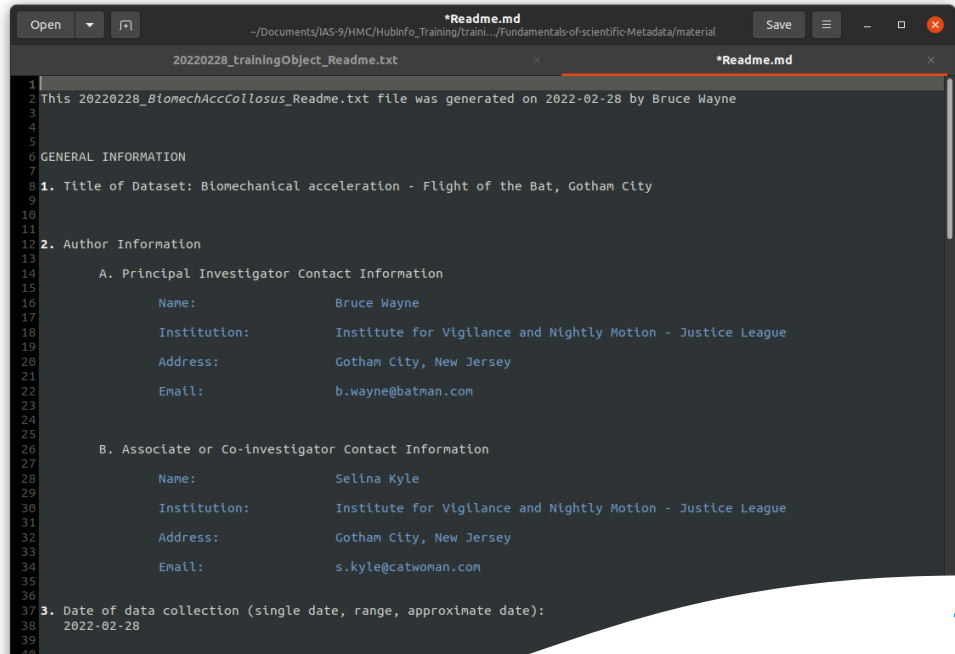
```
Open [ ] *Readme.md
~/Documents/IAS-9/HMC/HubInfo_Training/train.../Fundamentals-of-scientific-Metad... Save [ ] - [ ] [ ] [ ]
20220228_trainingObject_Readme.txt *Readme.md
1
2 This 20220228_BiomechAccCollosus_Readme.txt file was generated on 2022-02-28 by Bruce Wayne
3
4
5
6 GENERAL INFORMATION
7
8 1. Title of Dataset: Biomechanical acceleration - Flight of the Bat, Gotham City
9
10
11
12 2. Author Information
13
14   A. Principal Investigator Contact Information
15
16     Name:           Bruce Wayne
17
18     Institution:    Institute for Vigilance and Nightly Motion - Justice League
19
20     Address:        Gotham City, New Jersey
21
22     Email:          b.wayne@batman.com
23
24
25   B. Associate or Co-investigator Contact Information
26
27     Name:           Selina Kyle
28
29     Institution:    Institute for Vigilance and Nightly Motion - Justice League
30
31     Address:        Gotham City, New Jersey
32
33     Email:          s.kyle@catwoman.com
34
35
36
37 3. Date of data collection (single date, range, approximate date):
38    2022-02-28
39
40
```


Human readable metadata: README file

README files can be considered **a human-readable form of metadata recording.**

This information helps users understand **how to properly work with or reuse the data/code:**

- Name of the data document
- Title
- Author information
- The date the files were created
- Data fields/units
- PIDs
- Instrument info
- Sample info
- Date the files were last updated
- Version etc.



```
1 This 20220228_BiomechAccCollosus_Readme.txt file was generated on 2022-02-28 by Bruce Wayne
2
3
4
5
6 GENERAL INFORMATION
7
8 1. Title of Dataset: Biomechanical acceleration - Flight of the Bat, Gotham City
9
10
11
12 2. Author Information
13
14   A. Principal Investigator Contact Information
15
16       Name:           Bruce Wayne
17
18       Institution:      Institute for Vigilance and Nightly Motion - Justice League
19
20       Address:          Gotham City, New Jersey
21
22       Email:            b.wayne@batman.com
23
24
25   B. Associate or Co-investigator Contact Information
26
27       Name:             Selina Kyle
28
29       Institution:      Institute for Vigilance and Nightly Motion - Justice League
30
31       Address:          Gotham City, New Jersey
32
33       Email:            s.kyle@catwoman.com
34
35
36
37 3. Date of data collection (single date, range, approximate date):
38    2022-02-28
39
40
```

Even better – Readme style metadata

	A	B	C	D	E	
1	t	a				
2		0				
3		0.01				
4		0.04				
5		0.05				
6		0.06				
7		0.08				
8		0.1				
9		0.11				
10		0.14				
11		0.15				
12		0.16				
13		0.17				
14		0.2	0.558189	-0.1908045	-0.4121181	0
15		0.21	0.5764356	-0.1865862	-0.4162383	0
16		0.22	0.589581	-0.18639	-0.4258521	0
17		0.25	0.6049827	-0.1941399	-0.4243806	0
18		0.26	0.619992	-0.206991	-0.4192794	0
19		0.27	0.6320583	-0.2191554	-0.4092732	0
20		0.3	0.6392196	-0.2279844	-0.3975993	0
21		0.31	0.6465771	-0.2317122	-0.3908304	0
22		0.32	0.6583491	-0.2291616	-0.3950487	0
23		0.34	0.6725736	-0.2220984	-0.4050549	0



Flight of the bat

Results

txt

20220228_recordingData_Readme.txt

Institution: Institute for Vigilance and Nightly Motion - Justice League

Address: Gotham City, New Jersey

Email: b.wayne@batman.com

B. Associate or Co-Investigator Contact Information

Name: Selina Kyle

Institution: Institute for Vigilance and Nightly Motion - Justice League

Address: Gotham City, New Jersey

Email: s.kyle@catwoman.com

3. Date of data collection (single date, range, approximate date):
2022-02-28

Even better – Readme style metadata

	A	B			
1	t	ax			
2	0	0.3931			
3	0.01	0.3957			
4	0.04	0.4138			
5	0.05	0.4415481	-0.1		
6	0.06	0.4741173	-0.1		
7	0.08	0.5021739	-0.1		
8	0.1	0.5247369	-0.1		
9	0.11	0.5421987	-0.1		
10	0.14	0.5506353	-0.1		
11	0.15	0.5538726	-0.203067	-0.4057416	0
12	0.16	0.5534902	0.2025575	0.4056425	0
13	0.17	0.5527			
14	0.2	0.558			
15	0.21	0.5764			
16	0.22	0.589			
17	0.25	0.6049827	-0.1		
18	0.26	0.619992	-0		
19	0.27	0.6320583	-0.2191554	-0.4092732	0
20	0.3	0.6392196	-0.2279844	-0.3975993	0
21	0.31	0.6465771	-0.2317122	-0.3908304	0
22	0.32	0.6583491	-0.2291616	-0.3950487	0
23	0.34	0.6725736	-0.2220984	-0.4050549	0



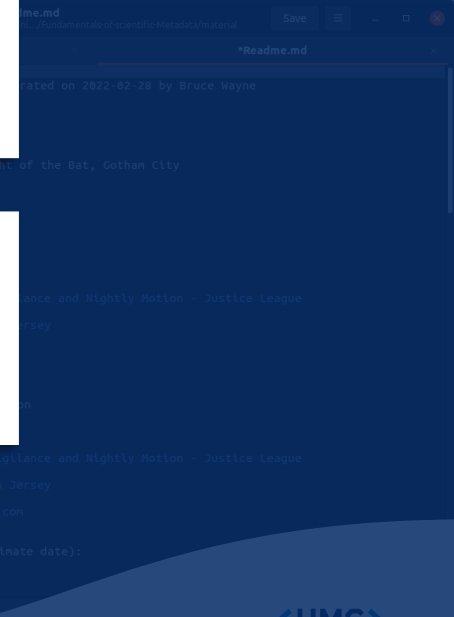
- documentation linked to the data
- locally searchable
- Readme file can be shared with the data
- increased readability



- unstructured
- subjective information
- only keyword search possible

dingData.csv

dingData_Readme.txt

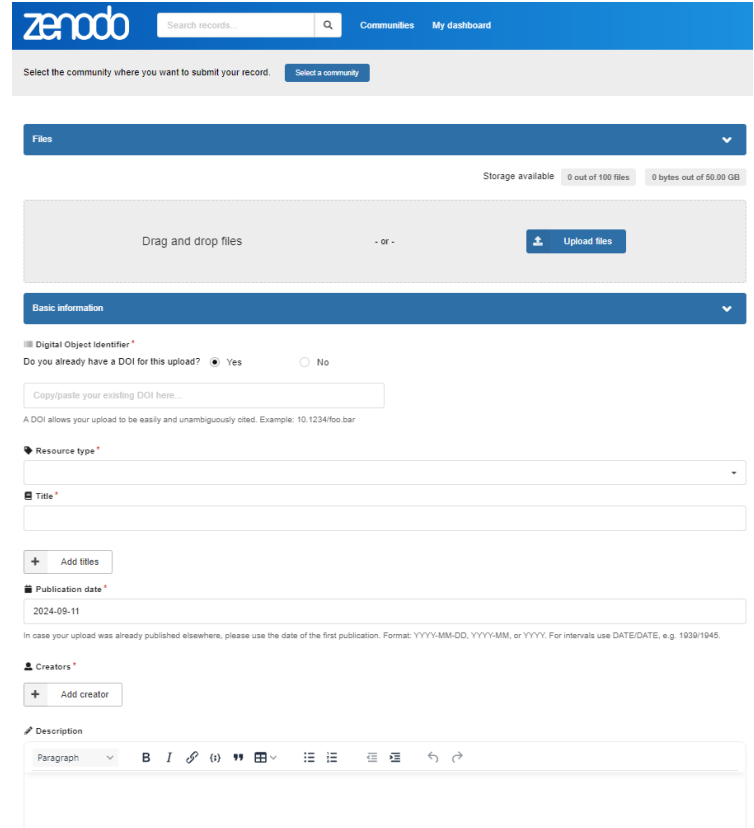


Metadata in repositories

While uploading your data to a repository, **you also enter the metadata into the system,**

just like you would when using Zenodo>>>

But there might be no field to define the variables in the dataset so we can **include a readme** file along with the dataset!



The screenshot shows the Zenodo upload interface. At the top, there is a search bar and navigation links for 'Communities' and 'My dashboard'. Below this, a section prompts the user to 'Select the community where you want to submit your record.' with a 'Select a community' button. The main area is divided into sections: 'Files', 'Basic information', and 'Description'. The 'Files' section shows storage availability (0 out of 100 files, 0 bytes out of 50.00 GB) and a 'Drag and drop files' area with an 'Upload files' button. The 'Basic information' section includes a 'Digital Object Identifier' field with radio buttons for 'Yes' (selected) and 'No', a text input for 'Copy/paste your existing DOI here...', and a note that 'A DOI allows your upload to be easily and unambiguously cited. Example: 10.1234/foo.bar'. Below this are fields for 'Resource type', 'Title', 'Publication date' (set to 2024-09-11), and 'Creators'. The 'Description' section features a rich text editor with a 'Paragraph' dropdown and various formatting options like bold, italic, link, and list.

Widely used formats for machine-readable metadata

XML

```
<example>
```

```
<superhero>Wonder Woman</superhero>
```

```
<publisher>DC Comics</publisher>
```

```
<identities>
```

```
<identity>Princess Diana</identity>
```

```
<identity>Diana Prince</identity>
```

```
</identities>
```

```
<pet>
```

```
<name>Jumpa</name>
```

```
<species>kangaroo</species>
```

```
</pet>
```

```
</example>
```

JSON

```
{
```

```
"superhero": "Wonder Woman",
```

```
"publisher": "DC Comics",
```

```
"identities": [
```

```
"Princess Diana",
```

```
"Diana Prince"
```

```
],
```

```
"pet": {
```

```
"name": "Jumpa",
```

```
"species": "kangaroo"
```

```
}
```

```
}
```

YAML

```
---
```

```
superhero: Wonder Woman
```

```
publisher: DC Comics
```

```
identities:
```

```
- Princess Diana
```

```
- Diana Prince
```

```
pet:
```

```
name: Jumpa
```

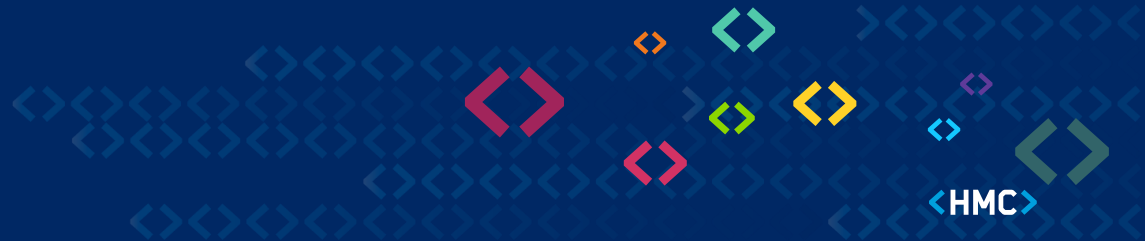
```
species: kangaroo
```

if you are interested in YAML,
also see <https://yaml.org/>

Questions?



Metadata Standard and Schema



Metadata Standards

This is a photo of my morning coffee >>>

- My phone also records the **metadata** attached to this photo.
- **Exchangeable image file format (EXIF)**
 - is most **widely used metadata standard** for images.
 - is embedded in image files by **digital cameras and smartphones**.
- These standards **ensure that metadata is structured** and used consistently across different platforms, software, and devices.



Metadata Standards

They are all in EXIF Metadata Standard:

```
plaintext Copy code
File Name: sunset_beach.jpg
File Type: JPEG
Make: Canon
Model: Canon EOS 5D Mark IV
Date/Time Original: 2024:08:25 19:05:12
Exposure Time: 1/400 sec
F-Number: f/5.6
ISO: 100
Focal Length: 24 mm
Flash: Off
GPS Latitude: 34.0195° N
GPS Longitude: 118.4912° W
Software: Adobe Photoshop 2024
Image Size: 6720x4480
White Balance: Auto
Metering Mode: Multi-segment
Orientation: Horizontal (normal)
Exposure Program: Aperture priority
```

```
plaintext Copy code
File Name: mountain_hike.jpg
File Type: JPEG
Make: Google
Model: Pixel 6 Pro
Date/Time Original: 2024:09:02 09:45:50
Exposure Time: 1/500 sec
F-Number: f/1.9
ISO: 50
Focal Length: 5.8 mm
Flash: Off
GPS Latitude: 39.7392° N
GPS Longitude: 104.9903° W
Software: Android 13
Image Size: 4032x3024
White Balance: Auto
Metering Mode: Center-weighted average
Orientation: Rotate 90 CW
Exposure Program: Auto
```

```
plaintext Copy code
File Name: morning_coffee.jpg
File Type: JPEG
Make: Apple
Model: iPhone 13
Date/Time Original: 2024:09:04 08:32:15
Exposure Time: 1/200 sec
F-Number: f/2.2
ISO: 40
Focal Length: 4.2 mm
Flash: Off, Did not fire
GPS Latitude: 37.7749° N
GPS Longitude: 122.4194° W
Software: iOS 17.6.1
Image Size: 3840x2160
White Balance: Auto
Metering Mode: Multi-segment
Orientation: Rotate 90 CW
Exposure Program: Program AE
```

Metadata Standards

XMP

```
plaintext Copy code
File Name: sunset_beach.jpg
File Type: JPEG
Creator: John Doe
Camera Model: Canon EOS 5D Mark IV
Lens: 24mm f/5.6
Date Created: 2024-08-25T19:05:12-07:00
Exposure: 1/400 sec
ISO: 100
Focal Length: 24 mm
Flash: Off
Scene Type: Directly Photographed
Software: Adobe Photoshop 2024
Image Size: 6720x4480
Location: GPS Latitude: 34.0195° N, GPS Longitude:
Copyright: © 2024 John Doe
Title: Sunset Beach
Description: A beautiful sunset at the beach with g
Keywords: sunset, beach, landscape, water, nature
Editing History: Edited: 2024-08-26 10:30:00
```

IPTC

```
plaintext Copy code
File Name: mountain_hike.jpg
File Type: JPEG
Creator: Jane Smith
Title: Mountain Adventure
Description: A breathtaking view from the peak of t
Keywords: mountain, hiking, nature, adventure
Date Created: 2024:09:02
Location: Rocky Mountains, Colorado, USA
Copyright Notice: © 2024 Jane Smith
Contact Info: janesmith@example.com
```

EXIF

```
plaintext Copy code
File Name: morning_coffee.jpg
File Type: JPEG
Make: Apple
Model: iPhone 13
Date/Time Original: 2024:09:04 08:32:15
Exposure Time: 1/200 sec
F-Number: f/2.2
ISO: 40
Focal Length: 4.2 mm
Flash: Off, Did not fire
GPS Latitude: 37.7749° N
GPS Longitude: 122.4194° W
Software: iOS 17.6.1
Image Size: 3840x2160
White Balance: Auto
Metering Mode: Multi-segment
Orientation: Rotate 90 CW
Exposure Program: Program AE
```

The Dublin Core

One of the best-known, generic, and widely used metadata standard for **online resources**, is the Dublin Core.

The Dublin Core was developed by a consortium of researchers, librarians, and web technologists in 1995, to address the need for a **unified description of web resources**.

Creator

Contributor

Publisher

Title

Date

Language

Format

Subject

Description

Identifier

Relation

[1] <https://www.dublincore.org/resources/metadata-basics/>

[2] <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/#section-3>

[3] <https://www.dublincore.org/about/>

[4] <https://www.iso.org/standard/71339.html>

https://www.science.org/

Dublin Core Elements



```
    /moatframe.js"></script>
    <meta name="pbContext" content=";issue:issue:doi\ :10.1126/science.2023.379.issue-6
ring:Publication Websites;pageGroup:string:Publication Pages">
    <link rel="schema.DC" href="http://purl.org/DC/elements/1.0/">
    <meta name="citation_journal_title" content="Science">
    <meta name="dc.Title" content="USB1 is a miRNA deadenylase that regulates hematopo
    <meta name="dc.Creator" content="Ho-Chang Jeong">
    <meta name="dc.Creator" content="Siddharth Shukla">
    <meta name="dc.Creator" content="Wilson Chun Fok">
    <meta name="dc.Creator" content="Thao Ngoc Huynh">
    <meta name="dc.Creator" content="Luis Francisco Zirnberger Batista">
    <meta name="dc.Creator" content="Roy Parker">
    <meta name="dc.Description" content="Mutations in the 3' to 5' RNA exonuclease USB
    U6 small nuclear RNA maturation, the molecular m...">
    <meta name="Description" content="Mutations in the 3' to 5' RNA exonuclease USB1 c
    small nuclear RNA maturation, the molecular m...">
    <meta name="dc.Publisher" content="American Association for the Advancement of Sci
    <meta name="dc.Date" scheme="WTN8601" content="2023-03-03">
    <meta name="dc.Type" content="research-article">
    <meta name="dc.Format" content="text/HTML">
    <meta name="dc.Identifier" scheme="publisher-id" content="abj8379">
    <meta name="dc.Identifier" scheme="doi" content="10.1126/science.abj8379">
    <meta name="dc.Language" content="EN">
```

html.pb-page.js > head > script

Filter Styles

:hov .cls +



Layout

Computed

Changes

Metadata schemas

Metadata schemas define the structure and fields for metadata, often built on a standard.

Key Difference

A standard defines **what metadata to include**:

- Dublin Core: Includes elements like title, creator, subject, date, description, etc,
- EXIF Metadata Schema: Includes fields like File Name, File Type, Model, Date/Time, etc.

A schema defines **how the metadata is structured and formatted**:

- Relationships between fields,
- Defining required or optional fields,
- Date: YYYY-MM-DD (ISO 8601 format),
- Name: Last Name, First Name,
- Keywords: A list separated by commas or semicolons.

Zenodo new upload form - **schemas** define the structure and fields for metadata

Files >

Storage available 0 out of 100 files 0 bytes out of 50.00 GB

Drag and drop files

- or -

Upload files

Basic information >

Digital Object Identifier *

Do you already have a DOI for this upload? Yes No

Copy/paste your existing DOI here...

A DOI allows your upload to be easily and unambiguously cited. Example: 10.1234/foo.bar

Resource type *

Title *

+ Add titles

Publication date *

2024-11-24

In case your upload was already published elsewhere, please use the date of the first publication. Format: YYYY-MM-DD, YYYY-MM, or YYYY. For intervals use DATE/DATE, e.g. 1939/1945.

Creators *

+ Add creator

Description

asterisks indicate mandatory entries

controlled list

entry format

Human-readable input (form fields) translates into machine-readable metadata

Files >

Storage available 0 out of 100 files 0 bytes out of 50.00 GB

Drag and drop files - or -

Basic information >

Digital Object Identifier *
Do you already have a DOI for this upload? Yes No

A DOI allows your upload to be easily and unambiguously cited. Example: 10.1234/foo.bar

Resource type *

Title *

Publication date *

In case your upload was already published elsewhere, please use the date of the first publication. Format: YYYY-MM-DD, YYYY-MM, or YYYY. For intervals use DATE/DATE, e.g. 1939/1945.

Creators *

Description

newdatasubmission.json

```
{  
  "Files": "metadatabeginners.pptx",  
  "Resource type": "Presentation",  
  "Title": "Metadata for Beginners",  
  "Publication date": "2024-11-24",  
  "Creators": "Dr. Özlem ÖZKAN"  
}
```


Domain-relevant community standards

Data should meet **domain-relevant community standards!**

Many disciplines have created

- metadata standards for describing data,
- created lists of recommended file formats etc.

Keeping in line with these standards will lead new data out into the ecosystem of data that is easy and suitable for others to reuse.

Interoperability: NASA's Mars Climate Orbiter Mishap

The failure of NASA's Mars Climate Orbiter in 1999

- The engineering team at Lockheed Martin used **English units of measurement (pounds-force)**,
- while the NASA team expected data in **metric units (newtons)**.

1 Newton \approx 0.224809 pounds-force (lbf)

Consequence:

The orbiter entered the atmosphere at a much lower altitude than intended, leading to **its destruction** by atmospheric stresses and heat.

Loss: \$327.6 million



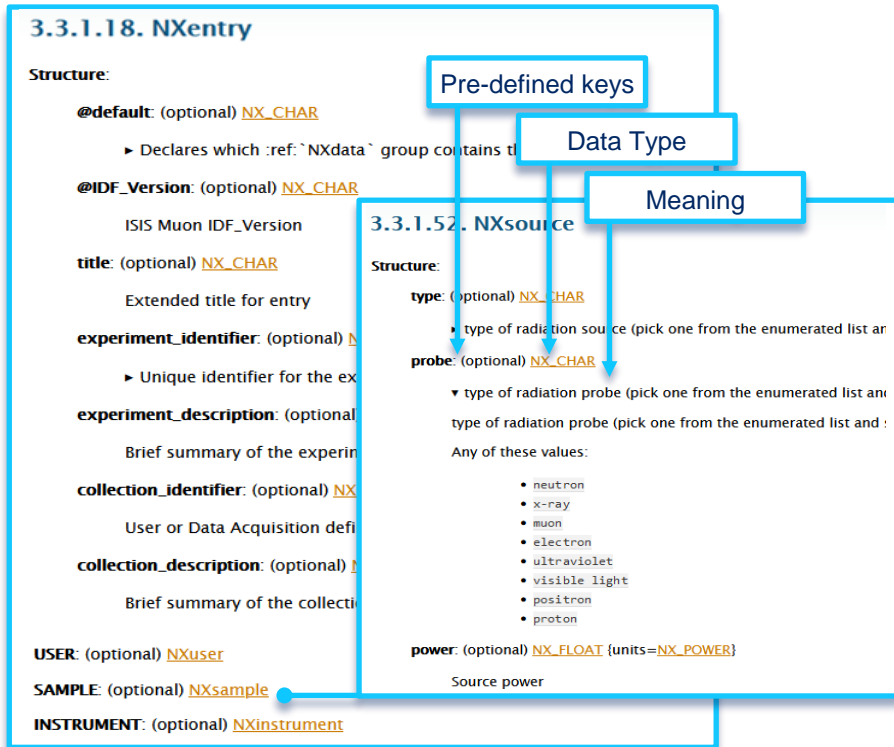
Remember the Mars Climate Orbiter incident from 1999?

What about research field matter?

NeXus is an **open community standard!**

[the **NeXus International Advisory Committee (NIAC)** (since 1994)]

It functions both as a **data format** and a **metadata standard** particularly in **neutron, X-ray, and muon research.**



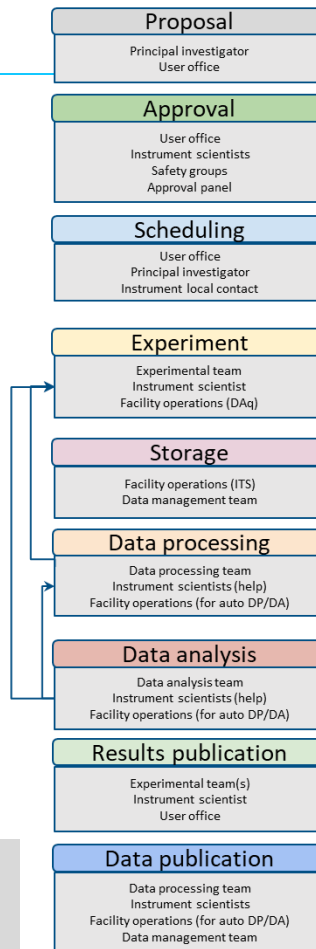
https://manual.nexusformat.org/classes/base_classes/NXsource.html

What about research field matter?

ExPaNDS project report (2022),

- defined the **necessary metadata elements** for Photon and Neutron (PaN) facilities >>>

(these elements are aligned with other open standards, like **NeXus**)



P1: essential
P2: important
P3: useful

PI/Main proposer	P1	FA
Co-investigators	P1	FA
Instrument requested	P1	F
Funding source	P2	F
Sample description	P1	F
Proposed experimental conditions (safety conditions)	P1	F
Experiment description	P3	F
Prior art (related publications, proposals)	P2	F
Facility information	P1	F
Proposal identifier	P1	F
[Approval panel]	P3	/
Sample safety assessment	P2	/
Allocated day & time on instrument	P2	FA
Scheduled visiting experimental team	P2	FA
Safety Training data	P3	/
Detailed experimental planning	P2	F
Sample preparation	P2	FR
[Sample reception]	P3	/
Visiting experimental team (user id)	P1	FA
Experiment date	P1	FA
Sample information	P1	FR
Instrument information	P1	FR
Calibration information	P1	FR
xperimental planning	P2	FR
Environmental parameters	P2	FR
Laboratory notebook	P2	FR
Instrument scientist	P2	F
[Experimental report]	P3	R
Persistent Identifiers (PIDs)	P1	FA
Preservation description information	P1	AR
Dataset information	P1	F
File identifier	P2	AR
[Representation information]	P3	IR
[Instrument parameters]	P3	FR
Processing team (user ID)	P2	AIR
Original data	P1	IR
Data format (after processing)	P1	IR
Dataset information	P2	AIR
Processing information	P1	R
Software package information	P1	R
Analysis team (user id)	P2	AIR
Original data	P1	IR
Software package information	P1	IR
Dependence tracking and workflow	P2	R
Data formats (after analysis)	P1	IR
Dataset information	P2	IR
File identifier	P1	AIR
[Instrument parameters]	P3	IR
[Calibration information]	P3	IR
Authors / Coauthors (user ID)	P1	FA
Proposal information	P1	FA
Publication information	P1	F
persistent identifier (PID)	P1	F
[Supplementary data information]	P3	F
Resource identity	P1	FI
Related resource	P2	F
Creator	P1	F
Contributor	P2	F
Title	P1	F
Publisher	P1	FI
Publication year	P1	FI
License	P1	IR
Release date	P1	IR

Metadata for nuclear physics experiments: NAPMIX

<https://www.oscars-project.eu/projects/napmix-nuclear-astro-and-particle-metadata-integration-experiments>

The NAPMIX project emerged to address **a significant gap in nuclear physics**: the lack of a unified metadata schema necessary for achieving FAIR datasets. This challenge extends to the fields of **particle and astro-particle physics**, highlighting the need for a collaborative European effort to **create a common metadata schema with user-friendly infrastructure**. By integrating expertise across these domains, NAPMIX aims to enhance data management practices.



The authors acknowledge the OSCARS project, which has received funding from the European Commission's Horizon Europe Research and Innovation programme under grant agreement No. 101129751

Funded by:



Funded by the European Union

Questions?



fairsharing.org



JSON

step-by-step



The JSON object



john.json

```
{  
  "name": "John",  
  "age": 27,  
  "employed": true,  
  "hasCar": null,  
  "parents": ["Anna", "Michael"],  
  "pet": {  
    "name": "Brutus",  
    "species": "dog",  
    "age": 7  
  }  
}
```

The JSON object

 john.json

```
{  
  "name": "John",  
  "age": 27,  
  "employed": true,  
  "hasCar": null,  
  "parents": ["Anna", "Michael"],  
  "pet": {  
    "name": "Brutus",  
    "species": "dog",  
    "age": 7  
  }  
}
```

Objects are enclosed in curly braces

The JSON object

 john.json

```
{  
  "name": "John",  
  "age": 27,  
  "employed": true,  
  "hasCar": null,  
  "parents": ["Anna", "Michael"],  
  "pet": {  
    "name": "Brutus",  
    "species": "dog",  
    "age": 7  
  }  
}
```

Information is
stored in
"key": "value"
pairs

The JSON object

 john.json

```
{  
  "name": "John",  
  "age": 27,  
  "employed": true,  
  "hasCar": null,  
  "parents": ["Anna", "Michael"],  
  "pet": {  
    "name": "Brutus",  
    "species": "dog",  
    "age": 7  
  }  
}
```

Information is stored in
"key": "value"
pairs

keys are of
datatype **string**

The JSON object

 john.json

```
{
  "name": "John",
  "age": 27,
  "employed": true,
  "hasCar": null,
  "parents": ["Anna", "Michael"],
  "pet": {
    "name": "Brutus",
    "species": "dog",
    "age": 7
  }
}
```

values **must be**
one of the following
data types:

- string
- number
- boolean
- null
- array
- object

The JSON object

 john.json

```
{
  "name": "John",
  "age": 27,
  "employed": true,
  "hasCar": null,
  "parents": ["Anna", "Michael"],
  "pet": {
    "name": "Brutus",
    "species": "dog",
    "age": 7
  }
}
```

values **must be**
one of the following
data types:

- string**
- number
- boolean
- null
- array
- object

The JSON object



john.json



strings are any kind of characters enclosed in “ “

- “word”
- “This is also a string.”
- “7 bananas”

```
{  
  "name": "John",  
  "age": 27,  
  "employed": true,  
  "hasCar": null,  
  "parents": ["Anna", "Michael"],  
  "pet": {  
    "name": "Brutus",  
    "species": "dog",  
    "age": 7  
  }  
}
```

boolean
null
array
object

It be
following

The JSON object

 john.json

```
{
  "name": "John",
  "age": 27,
  "employed": true,
  "hasCar": null,
  "parents": ["Anna", "Michael"],
  "pet": {
    "name": "Brutus",
    "species": "dog",
    "age": 7
  }
}
```

values **must be**
one of the following
data types:

- string
- number**
- boolean
- null
- array
- object

The JSON object



john.json



numbers can be:

- integers (e.g. 42)
- floats (e.g. 0.0005)

```
{  
  "name": "John",  
  "age": 27,  
  "employed": true,  
  "hasCar": null,  
  "parents": ["Anna", "Michael"],  
  "pet": {  
    "name": "Brutus",  
    "species": "dog",  
    "age": 7  
  }  
}
```

boolean
null
array
object

It can be
following

The JSON object



john.json

```
{
  "name": "John",
  "age": 27,
  "employed": true,
  "hasCar": null,
  "parents": ["Anna", "Michael"],
  "pet": {
    "name": "Brutus",
    "species": "dog",
    "age": 7
  }
}
```

values must be
one of the following
data types:

- string
- number
- boolean**
- null
- array
- object

The JSON object



john.json



a **boolean** has one of two possible values

- true / false
- 1 / 0

```
{  
  "name": "John",  
  "age": 27,  
  "employed": true,  
  "hasCar": null,  
  "parents": ["Anna", "Michael"],  
  "pet": {  
    "name": "Brutus",  
    "species": "dog",  
    "age": 7  
  }  
}
```

boolean
null
array
object

it be
following

The JSON object

 john.json

```
{
  "name": "John",
  "age": 27,
  "employed": true,
  "hasCar": null,
  "parents": ["Anna", "Michael"],
  "pet": {
    "name": "Brutus",
    "species": "dog",
    "age": 7
  }
}
```

values **must be**
one of the following
data types:

string
number
boolean
null
array
object

The JSON object



john.json



null can only have the value NULL.
The variable of data type **null** has
no value assigned to it.

```
{  
  "name": "John",  
  "age": 27,  
  "employed": true,  
  "hasCar": null,  
  "parents": ["Anna", "Michael"],  
  "pet": {  
    "name": "Brutus",  
    "species": "dog",  
    "age": 7  
  }  
}
```

boolean
null
array
object

It be
following

The JSON object

 john.json

```
{
  "name": "John",
  "age": 27,
  "employed": true,
  "hasCar": null,
  "parents": ["Anna", "Michael"],
  "pet": {
    "name": "Brutus",
    "species": "dog",
    "age": 7
  }
}
```

values **must be**
one of the following
data types:

- string
- number
- boolean
- null
- array**
- object

The JSON object



john.json



An **array** is a collection of elements.
Can be understood as a list.

- ["Bibi", "Tina"]
- [1,2,3]
- ["some string", 0.5, true]

```
{  
  "name": "John",  
  "age": 27,  
  "employed": true,  
  "hasCar": null,  
  "parents": ["Anna", "Michael"],  
  "pet": {  
    "name": "Brutus",  
    "species": "dog",  
    "age": 7  
  }  
}
```

boolean
null
array
object

It be
following

The JSON object

 john.json

```
{  
  "name": "John",  
  "age": 27,  
  "employed": true,  
  "hasCar": null,  
  "parents": ["Anna", "Michael"],  
  "pet": {  
    "name": "Brutus",  
    "species": "dog",  
    "age": 7  
  }  
}
```

values **must be**
one of the following
data types:

- string
- number
- boolean
- null
- array
- object**

The JSON object



john.json



An object contains key/value pairs, separated by commas and is enclosed by `{ }`

```
{  
  "name": "Bill",  
  "jobTitle": "Postdoc",  
  "city": "New York",  
  "age": 36  
}
```

```
{  
  "name": "John",  
  "age": 27,  
  "employed": true,  
  "hasCar": null,  
  "parents": ["Ann", "John"],  
  "pet": {  
    "name": "Brutus",  
    "species": "dog",  
    "age": 7  
  }  
}
```

```
  "name": "Brutus",  
  "species": "dog",  
  "age": 7
```

```
  null  
  array  
  object
```

st be
following

The JSON object

 john.json

```
{  
  "name": "John",  
  "age": 27,  
  "employed": true,  
  "hasCar": null,  
  "parents": ["Anna", "Michael"],  
  "pet": {  
    "name": "Brutus",  
    "species": "dog",  
    "age": 7  
  }  
}
```

Data is
separated by
commas

The JSON object - indentation



john.json

```
{
  "name": "John",
  "age": 27,
  "employed": true,
  "hasCar": null,
  "parents": ["Anna", "Michael"],
  "pet": {
    "name": "Brutus",
    "species": "dog",
    "age": 7
  }
}
```

Structured metadata: JSON vs. XML

 john.json

```
{  
  "name": "John",  
  "age": 27,  
  "employed": true,  
  "hasCar": null,  
  "parents": ["Anna", "Michael"],  
  "pet": {  
    "name": "Brutus",  
    "species": "dog",  
    "age": 7  
  }  
}
```

 john.xml

```
<name>John</name>  
<age>27</age>  
<employed>TRUE</employed>  
<hasCar>NULL</hasCar>  
<parents>Anna</parents>  
<parents>Michael</parents>  
<pet>  
  <name>Brutus</name>  
  <species>dog</species>  
  <age>7</age>  
</pet>
```

Coding



Example:

```
{  
  "firstName": "value",  
  "ORCID": "value",  
  "favoriteCake": "value",  
  "hobbies": ["value", "value"],  
  "city": {  
    "name": "value",  
    "url": "value"  
  }  
}
```

Thanks

